

Markovkedjor

Markovkedjor beskriver hur ett system med flera olika tillstånd kan förändras.

Övergångsmatrisen

är en matris A , där elementet p_{ji} representerar sannolikheterna att gå från tillstånd $i \rightarrow j$.

Ursprungsvektorn

betecknas \vec{v}_0 och representerar systemets ursprungstillstånd.

Tillståndsvektorer

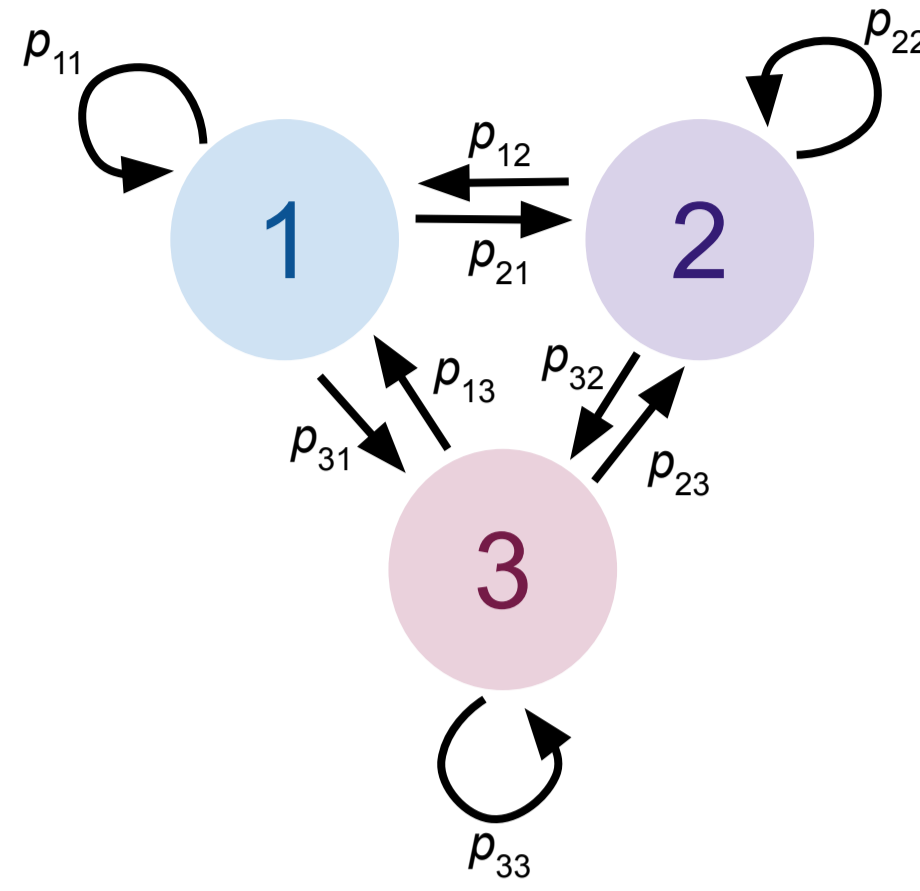
betecknas \vec{v}_i och representerar sannolikheterna för respektive tillstånd

$$A \cdot \vec{v}_0 = \vec{v}_1$$

$$\dots$$

$$A \cdot \vec{v}_i = \vec{v}_{i+1} = A^{i+1} \cdot \vec{v}_0$$

Detta ger sannolikheten för tillstånden efter $i+1$ upprepningar.



Ovan illustreras en markovkedja med tillstånden 1, 2 och 3, där p_{ji} är sannolikheten för en viss tillståndsövergång.

Nedan visas tillhörande övergångsmatris A .

$$A = \begin{matrix} \text{föregående tillstånd} & & & & \\ & 1 & 2 & 3 & \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{matrix} \text{kommande} \\ \text{tillstånd} \end{matrix} \end{matrix}$$

Lösningsmetod

Metoden bygger på att man utifrån det nuvarande ordet genererar ett nytt ord genom att multiplicera tillståndsvektorn med övergångsmatrisen, som i exemplet ovan. Man får då en vektor med sannolikheter för nästa ord. Med dessa sannolikheter slumpades nästa ord fram där sannolikheterna i vektorn användes som vikt.

	du	gamla	fria	fjällhöga	nord
du	0	1	1	0	0
gamla	1/3	0	0	0	0
fria	1/3	0	0	0	0
fjällhöga	1/3	0	0	0	0
nord	0	0	0	1	0

 $\Rightarrow \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/3 \\ 1/3 \\ 1/3 \\ 0 \end{bmatrix}$$

↑ Övergångsmatris ↑ Tillståndsvektor

Ord som avslutar en mening och påbörjar en mening är inte alltid direkt korrelerade, men det kan ändå finnas samband mellan dem. För att representera detta i övergångsmatrisen gjorde vi så att ord med skiljetecken mellan sig bidrar 80% mindre till den tillståndsövergången. Om en låt innehåller "Du, gamla du fria" skulle det vara mer sannolikt att "du" följs av "fria" än "gamla". Vi gjorde detta då vår datamängd var för liten för att ha med skiljetecken i övergångsmatrisen och vi ville ha en balans mellan grammatisk korrekthet och att låtarna höll sig till ett tema.

För att få mer sammanhängande resultat kan man kolla på sannolikheten att ett visst ord följer sekvenser på två eller fler ord. Datamängden vi använde var relativt liten och många tvåords-sekvenser är därav ovanliga eller till och med helt unika, vilket gör att låtar genererade med denna metod blir väldigt lika dataunderlaget.

Vårt projekt

Markovkedjor används ofta för musik- och textgenerering. Vi kombinerade idéerna för att skapa ett program som genererar sångtexter till Fysiksektionens sångbok.

Vårt program använder markovkedjor för att generera låtar. De är delvis sammanhängande, har liknande tema till sångboken, samt har passande antal stavelser för användarens valda melodi.

Resultat

Vi ska nu visa första stycket av en låt som programmet har genererat. Vi startade på ordet "matte" och valde melodin "Du gamla du fria":

**Matte är ensam överleva utan
spill tills kassan gått i glas det tar en
sup och vi dricker champis det som har defintion
anonyma människor var vi inte var
i mig en nubbe en ska jag kommer vi**

För jämförelse har vi också tagit oss friheten att skapa ett program som helt slumpmässigt genererar ord så att det framgår hur vårt program skiljer sig från slumpmässiga låtar. Vi valde att återigen ta första stycket från en låt som börjar på ordet "matte" och melodin "Du gamla du fria":

**Matte ene saft online snöra brevi
lön mekanik laphroig utom svalkar
lättad impotensbesvär presentera
helnkyterhet nektarin hell draperi
kräket tutigalen port vardagsrum fritt**

Av textstyckena framgår det att sångtexter producerade med markovkedjor varken är lika sammanhängande eller grammatiskt korrekta som texter skrivna av människor, men att de ändå presterar bättre inom respektive område än slumpmässigt genererade låttexter.

Sammanfattning

Markovkedjor används för att beskriva hur system i olika tillstånd kan förändras och används ibland för att generera texter eller melodier. Vi har skapat ett program som genererar låttexter utifrån Konglig Fysiks Sångbok.

Källor:

Anton, H. Rorres, C. (u.å). *Elementary Linear Algebra*. Fysiksektionen. (2022). *Sångboken*.