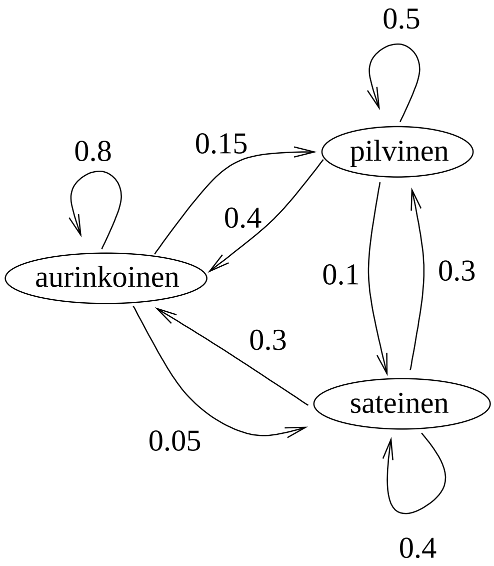
T-61.281 Luonnollisen kielen tilastollinen käsittely

Vastaukset 6, ti 2.3.2004, 8:30-10:00 Markov-ketjut ja kätketyt Markov-mallit, Versio 1.0

1. a) Kuvaan 1 on piirretty Turun säätila Markov-ketjuna.



Kuva 1: Säätila Markov-ketjuna

b) Lasketaan tilasekvenssin $\mathbf{S} = (S_3, S_2, S_1, S_1, S_1)$ todennäköisyys kun tiedetään, että lähdetään tilasta S_2. Haluamme siis laskea todennäköisyyttä

$$\begin{aligned}
P(\mathbf{S} \mid q_0 = S_2) &= P(q_1 = S_3, q_2 = S_2, q_3 = S_1, q_4 = S_1, q_5 = S_1 \mid q_0 = S_2) \\
&= P(q_1 = S_3 \mid q_0 = S_2) \cdot P(q_2 = S_2 \mid q_0 = S_2, q_1 = S_3) \\
&\quad \cdot P(q_3 = S_1 \mid q_0 = S_2, q_1 = S_3, q_2 = S_2) \\
&\quad \cdot P(q_4 = S_1 \mid q_0 = S_2, q_1 = S_3, q_2 = S_2, q_3 = S_1) \\
&\quad \cdot P(q_5 = S_1 \mid q_0 = S_2, q_1 = S_3, q_2 = S_2, q_3 = S_1, q_4 = S_1)
\end{aligned}$$

Ensimmäisen rivin todennäköisyys on jaettu tässä käyttäen monta kertaa todennäköisyyslaskun kaavaa $P(A, B|C) = P(A|C) \cdot P(B|A, C)$.

Sovelletaan ylläolevaan kaavaan Markov-oletusta, että nykyinen tila riippuu vain edellisestä tilasta, muttei sitä aikaisemmista:

$$\begin{aligned}
P(\mathbf{S} \mid q_0 = S_2) &= P(q_1 = S_3 \mid q_0 = S_2) \cdot P(q_2 = S_2 \mid q_1 = S_3) \\
&\quad \cdot P(q_3 = S_1 \mid q_2 = S_2) \cdot P(q_4 = S_1 \mid q_3 = S_1) \\
&\quad \cdot P(q_5 = S_1 \mid q_4 = S_1)
\end{aligned}$$

Tämä vastaa taulukoituja kertoimia a_{ij}:

$$\begin{aligned} P(\mathbf{S} \mid q_0 = S_2) &= a_{23} \cdot a_{32} \cdot a_{21} \cdot a_{11} \cdot a_{11} \\ &= 0.1 \cdot 0.3 \cdot 0.4 \cdot 0.8 \cdot 0.8 \\ &= 0.0077 \end{aligned}$$

c) Oletusarvo sille, kuinka kauan tilassa pysytään on

$$\begin{aligned} E(x) &= \int xP(x)dx \\ &= \sum_{n=1}^{\infty} n a_{ii}^n (1 - a_{ii}) \\ &= (1 - a_{ii}) \frac{a_{ii}}{(1 - a_{ii})^2} \\ &= \frac{a_{ii}}{1 - a_{ii}} \end{aligned}$$

Aurinkoisen päivän tapauksessa $a_{11} = 0.8$ eli saadaan $\frac{0.8}{0.2} = 4$ päivää. Tämä siis oli oletusarvo sille, kuinka kauan tilassa pysytään, eli mehän olimme ensimmäisenä päivänä jo valmiiksi aurinkoisessa tilassa ja lopullinen vastaus on $4 + 1 = 5$ päivää.

2. a) Tässä tehtävässä siis halutaa laskea tilan S_i todennäköisyys kolmantena päivänä kun tiedetään, että lähtöpäivänä oli aurinkoista ja seuraavan kolmen päivän lämpötilat $\mathbf{X} = (x_1, x_2, x_3) = (7°C, 3°C, -8°C)$. Merkitään kaikkia malliin liittyviä parametrejä $\lambda = \{\mathbf{A}, b_i(x)\}$.

$$\begin{aligned} & P(q_3 = S_i \mid q_0 = S_1, x_1, x_2, x_3, \lambda) \\ = & \frac{P(q_3 = S_i, x_1, x_2, x_3 \mid q_0 = S_1, \lambda)}{P(x_1, x_2, x_3 \mid q_0 = S_1, \lambda)} \\ = & \frac{P(q_3 = S_i, x_1, x_2, x_3 \mid q_0 = S_1, \lambda)}{\sum_{j=1}^{3} P(q_3 = S_j, x_1, x_2, x_3 \mid q_0 = S_1, \lambda)} \end{aligned}$$

Ylläolevan yhtälön pyörittämiseen käytettiin ensiksi todennäköisyyslaskun kaavaa $P(A|B,C) = \frac{P(A,B|C)}{P(B|C)}$. Sitten huomattiin että $P(A) = \sum_B P(A,B)$. Edellisen yhtällön nimittäjässä ja osoittajassa on samankaltainen termi. Kevennetään hieman merkintöjä ja esitetään sama asia funktion $\alpha_t(i)$ avulla:

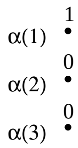
$$P(q_3 = S_i \mid q_0 = S_1, x_1, x_2, x_3, \lambda) = \frac{\alpha_3(i)}{\sum_{j=1}^{3} \alpha_3(j)}$$

Tarkastellaanpa, miten tämä eteenpäin-todennäköisyys $\alpha_t(i)$ oikein voidaan laskea.

Lähtöpäivä

Tiedetään, että lähtöpäivänä oli aurinkoista, joten voidaan asettaa

$$
\begin{aligned}
\alpha_0(1) &= 1\\
\alpha_0(2) &= 0\\
\alpha_0(3) &= 0
\end{aligned}
$$

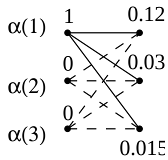


Eteenpäin-algoritmin hilan alustus

Ensimmäinen päivä

Edellisenä päivänä oli aurinkoista ja tänään on $7°C$ astetta lämmintä. Lasketaan kullekin tilalle transitiotodennäköisyydet aurinkoisesta ja kerrotaan ne emissiotodennäköisyyksillä, kun $x_1 > 5°C$.

$$
\begin{aligned}
\alpha_1(1) &= a_{11} \cdot b_1(x \geq 5°C) = 0.8 \cdot 0.15 = 0.120\\
\alpha_1(2) &= a_{12} \cdot b_2(x \geq 5°C) = 0.15 \cdot 0.2 = 0.030\\
\alpha_1(3) &= a_{13} \cdot b_3(x \geq 5°C) = 0.05 \cdot 0.3 = 0.015
\end{aligned}
$$

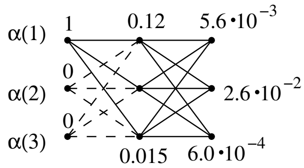


Hila ensimmäisen päivän jälkeen

Toinen päivä

Nyt emme tiedä, minkälainen sää edellisenä päivänä oikeasti oli. Summataan kaikki vaihtoehdot yhteen todennäköisyydellä painotettuna

$$
\begin{aligned}
\alpha_2(1) &= \sum_{j=1}^{3} \alpha_1(j) \cdot a_{j1} b_1(-5°C \leq x \leq 5°C)\\
&= (\alpha_1(1) \cdot a_{11} + \alpha_1(2) \cdot a_{21} + \alpha_1(3) \cdot a_{31})\\
&\quad \cdot b_1(-5°C \leq x \leq 5°C)\\
&= (0.8{\cdot}0.12{+}0.4{\cdot}0.03{+}0.3{\cdot}0.015) \cdot 0.05\\
&= 5.625 \cdot 10^{-3}\\
\alpha_2(2) &= \sum_{j=1}^{3} \alpha_1(j) \cdot a_{j2} b_2(-5°C \leq x \leq 5°C)\\
&= (0.15{\cdot}0.12{+}0.5{\cdot}0.03{+}0.3{\cdot}0.015) \cdot 0.7\\
&= 2.625 \cdot 10^{-2}\\
\alpha_2(3) &= \sum_{j=1}^{3} \alpha_1(j) \cdot a_{j3} b_3(-5°C \leq x \leq 5°C)\\
&= (0.05{\cdot}0.12{+}0.1{\cdot}0.03{+}0.4{\cdot}0.015) \cdot 0.4\\
&= 6.000 \cdot 10^{-4}
\end{aligned}
$$

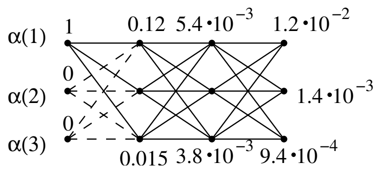


Hila toisen päivän jälkeen

Kolmas päivä

Jatketaan jo tutulla tavalla eteenpäin. $x_3 < -5°C$.

$$
\begin{aligned}
\alpha_3(1) &= \sum_{j=1}^{3} \alpha_2(j) \cdot a_{j1} b_1(-5°C \le x \le 5°C) \\
&= (\alpha_2(1) \cdot a_{11} + \alpha_2(2) \cdot a_{21} + \alpha_2(2) \cdot a_{31}) \\
&\quad \cdot b_1(x \le -5°C) \\
&= (0.8 \cdot 5.625 \cdot 10^{-3} + 0.4 \cdot 2.625 \cdot 10^{-2} \\
&\quad +0.3 \cdot 6.000 \cdot 10^{-4}) \cdot 0.8 \\
&= 1.2144 \cdot 10^{-2} \\
\alpha_3(2) &= \sum_{j=1}^{3} \alpha_2(j) \cdot a_{j2} b_2(-5°C \le x \le 5°C) \\
&= (0.15 \cdot 5.625 \cdot 10^{-3} + 0.5 \cdot 2.625 \cdot 10^{-2} \\
&\quad +0.3 \cdot 6.000 \cdot 10^{-4}) \cdot 0.1 \\
&= 1.4149 \cdot 10^{-3} \\
\alpha_3(3) &= \sum_{j=1}^{3} \alpha_2(j) \cdot a_{j3} b_3(-5°C \le x \le 5°C) \\
&= (0.05 \cdot 5.625 \cdot 10^{-3} + 0.1 \cdot 2.652 \cdot 10^{-2} \\
&\quad +0.4 \cdot 6.000 \cdot 10^{-4}) \cdot 0.3 \\
&= 9.4387 \cdot 10^{-4}
\end{aligned}
$$



Hila viimeisenä päivänä

Nyt olemme saaneet lasketuksi tehtävän ratkaisuun tarvittavat suureet. Sijoitetaan luvut kaavaan:

$$
\begin{aligned}
P(q_3 = S_1 \mid q_0 = S_1, x_1, x_2, x_3, \lambda) &= \frac{\alpha_3(1)}{\sum_{i=j}^{3} \alpha_3(j)} \\
&= \frac{1.2144 \cdot 10^{-2}}{1.2144 \cdot 10^{-2} + 1.4149 \cdot 10^{-3} + 9.4387 \cdot 10^{-4}} \\
&= 0.8874
\end{aligned}
$$

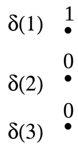
Palatessa on siis 89 % todennäköisyydellä aurinkoista. Samalla tavalla voidaan laskea pilvisen sään todennäköisyys 10 % ja sateisen sään todennäköisyys 1 %.

b) Halusimme vielä arvata, mikä oli todennäköisin sääsekvenssi poissa ollessamme. Tämän voimme hoitaa Viterbi-algoritmilla, joka on hyvin samankaltainen eteenpäin-algoritmin kanssa. Nyt vain joka tilassa lasketaan todennäköisyys parhaan reitin kautta sen sijaan että summattaisiin kaikkien reittejen yli.

Viterbi: Lähtöpäivä

Alustetaan hila samalla tavalla kuin eteenpäin-algoritmissa:

$$\begin{aligned}\delta_0(1) &= 1\\ \delta_0(2) &= 0\\ \delta_0(3) &= 0\end{aligned}$$

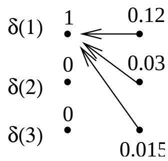


Viterbi-algoritmin hilan alustus

Viterbi: Ensimmäinen päivä

Paras polku kuhunkin tilaan tulee aurinkoisesta. Lähtöpäivän säähän tiedettiin varmasti, joten muiden säätilojen todennäköisyys on nolla. Merkitään kuhunkin tilaan, mistä todennäköisin polku tulee ($\psi_1(i)$). Laskutoimitukset sinällään ovat vielä aivan samat kuin eteenpäin-algoritmissakin.

$$\begin{aligned}\delta_1(1) &= a_{11} \cdot b_1(x \geq 5°C) = 0.8 \cdot 0.15 = 0.120\\ \delta_1(2) &= a_{12} \cdot b_2(x \geq 5°C) = 0.15 \cdot 0.2 = 0.030\\ \delta_1(3) &= a_{13} \cdot b_3(x \geq 5°C) = 0.05 \cdot 0.3 = 0.015\\ \psi_1(1) &= 1\\ \psi_1(2) &= 1\\ \psi_1(3) &= 1\end{aligned}$$



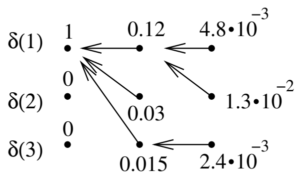
Viterbi-haku ensimmäisen päivän jälkeen

Viterbi: Toinen päivä

Valitaan tilaan tulevasta reitistä todennäköisin:

$$\begin{aligned}\delta_2(1) &= \max_j \left(\delta_1(j) \cdot a_{j1} b_1(-5°C \leq x \leq 5°C)\right)\\ &= \max\left(\delta_1(1) \cdot a_{11}, \delta_1(2) \cdot a_{21}, \delta_1(3) \cdot a_{31}\right)\\ &\quad \cdot b_1(-5°C \leq x \leq 5°C)\\ &= \max\left(0.8 \cdot 0.12, 0.4 \cdot 0.03, 0.3 \cdot 0.015\right)\\ &\quad \cdot 0.05\\ &= \max\left(9.6 \cdot 10^{-2}, 1.2 \cdot 10^{-2}, 4.5 \cdot 10^{-3}\right)\\ &\quad \cdot 0.05\\ &= 9.6 \cdot 10^{-2} \cdot 0.05 = 4.8 \cdot 10^{-3}\\ \psi_2(1) &= \operatorname*{argmax}_j(\delta_1(j) \cdot a_{j1} b_1(-5°C \leq x \leq 5°C))\\ &= 1\end{aligned}$$

$$
\begin{aligned}
\delta_2(2) &= \max_j \left(\delta_1(j) \cdot a_{j2} b_2(-5°C \leq x \leq 5°C)\right) \\
&= \max(0.15 \cdot 0.12, 0.5 \cdot 0.03, 0.3 \cdot 0.015) \\
&\quad \cdot 0.7 \\
&= \max\left(1.8 \cdot 10^{-2}, 1.5 \cdot 10^{-2}, 4.5 \cdot 10^{-3}\right) \\
&\quad \cdot 0.7 \\
&= 1.8 \cdot 10^{-2} \cdot 0.7 = 1.26 \cdot 10^{-2} \\
\psi_2(2) &= \operatorname*{argmax}_j(\delta_1(j) \cdot a_{j2} b_2(-5°C \leq x \leq 5°C)) \\
&= 1 \\
\delta_2(3) &= \max_j \left(\delta_1(j) \cdot a_{j3} b_3(-5°C \leq x \leq 5°C)\right) \\
&= \max(0.05 \cdot 0.12, 0.1 \cdot 0.03, 0.4 \cdot 0.015) \\
&\quad \cdot 0.4 \\
&= \max\left(6.0 \cdot 10^{-3}, 3.0 \cdot 10^{-3}, 6.0 \cdot 10^{-3}\right) \\
&\quad \cdot 0.4 \\
&= 6.0 \cdot 10^{-3} \cdot 0.4 = 2.4 \cdot 10^{-3} \\
\psi_2(3) &= \operatorname*{argmax}_j(\delta_1(j) \cdot a_{j3} b_2(-5°C \leq x \leq 5°C)) \\
&= 3
\end{aligned}
$$



Hila toisen päivän jälkeen

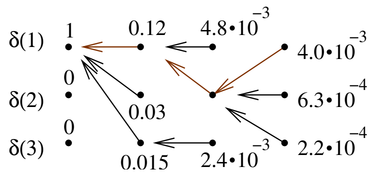
Laskettaessa $\psi_2(3)$ huomataan, että kahdesta paikkaa (tilat 1 ja 3) päästään yhtä todennäköisesti tilaan 3. Tässä tapauksessa paras paluureitti ratkaistiin arpomalla näiden kahden välillä ja tulokseksi saatiin 3.

Viterbi: Kolmas päivä

Valitaan tilaan tulevasta reitistä todennäköisin:

$$
\begin{aligned}
\delta_3(1) &= \max_j \left(\delta_2(j) \cdot a_{j1} b_1(x \leq -5°C)\right) \\
&= \max(\delta_2(1) \cdot a_{11}, \delta_2(2) \cdot a_{21}, \delta_2(3) \cdot a_{31}) \\
&\quad \cdot b_1(x \leq -5°C) \\
&= \max(0.8 \cdot 4.8 \cdot 10^{-3}, 0.4 \cdot 1.26 \cdot 10^{-2}, \\
&\quad 0.3 \cdot 2.4 \cdot 10^{-3}) \cdot 0.8 \\
&= \max\left(3.8 \cdot 10^{-3}, 5.0 \cdot 10^{-3}, 7.2 \cdot 10^{-4}\right) \\
&\quad \cdot 0.8 \\
&= 4.0 \cdot 10^{-3} \\
\psi_3(1) &= \operatorname*{argmax}_j(\delta_2(j) \cdot a_{j1} b_1(x \leq 5°C)) = 2
\end{aligned}
$$

$$
\begin{aligned}
\delta_3(2) &= \max_j \left(\delta_2(j) \cdot a_{j2} b_2(x \leq -5°C)\right) \\
&= \max(0.15 \cdot 4.8 \cdot 10^{-3}, 0.5 \cdot 1.26 \cdot 10^{-2}, \\
&\quad 0.3 \cdot 2.4 \cdot 10^{-3}) \cdot 0.1 \\
&= \max\left(7.2 \cdot 10^{-4}, 6.3 \cdot 10^{-3}, 7.2 \cdot 10^{-4}\right) \\
&\quad \cdot 0.1 \\
&= 6.3 \cdot 10^{-3} \cdot 0.1 = 6.3 \cdot 10^{-4} \\
\psi_2(2) &= \operatorname*{argmax}_j(\delta_1(j) \cdot a_{j2} b_2(x \leq -5°C)) = 2 \\
\delta_3(3) &= \max_j \left(\delta_2(j) \cdot a_{j3} b_3(x \leq -5°C)\right) \\
&= \max(0.05 \cdot 4.8 \cdot 10^{-3}, 0.1 \cdot 1.26 \cdot 10^{-2}, \\
&\quad 0.4 \cdot 2.4 \cdot 10^{-3}) \cdot 0.3 \\
&= \max\left(2.4 \cdot 10^{-4}, 1.3 \cdot 10^{-3}, 7.2 \cdot 10^{-4}\right) \\
&\quad \cdot 0.3 \\
&= 2.16 \cdot 10^{-4} \\
\psi_3(3) &= \operatorname*{argmax}_j(\delta_1(j) \cdot a_{j3} b_2(x \leq -5°C)) = 2
\end{aligned}
$$



Viterbi-haun hila viimeisenä päivänä

Valmiista hilasta voidaan hakea todennäköisin tilasekvenssi: aloitetaan kaikkein todennäköisimmästä lopputilasta ja seurataan nuolia alkuun päin. Nyt siis näyttäisi siltä, että aurinkoisen lähtöpäivän jälkeen Turussa on ollut aurinkoista, pilvistä ja taasen aurinkoista.

Yhteenveto: mitä eroa eteenpäin-algoritmilla ja Viterbi-haulla

Eteenpäin-algoritmi hakee oikean todennäköisyyden kullekin tilasekvenssille. Sen avulla ei kuitenkaan pysty hakemaan parasta polkua hilasta.

Viterbi-haulla tilatodennäköisyydet ovat vain approksimaatioita. Tätä algorimia kuitenkin käytetään paljon juuri sen takia, että sillä pysytytään löytämään paras polku.

Laskennallisesti algoritmit ovat yhtä raskaita, eteenpäin-algoritmin summaus on vain Viterbi-haussa vaihtunut maksimoinniksi.