Rozpoznávání s markovskými modely

Václav Hlaváč

České vysoké učení technické v Praze

Český institut informatiky, robotiky a kybernetiky

160 00 Praha 6, Jugoslávských partyzánů 1580/3

http://people.ciirc.cvut.cz/hlavac, vaclav.hlavac@cvut.cz

Poděkování: M.I. Schlesinger

Osnova přednášky:

- Motivace, použití.
- Stochastické konečné automaty.
- Markovský statistický model.
- Tři nejčastější úlohy se skrytými markovskými modely (rozpoznávání, hledání nejpravděpodobnější posloupnosti stavů, empirické učení markovských statistických modelů).

Klasifikace závislá na kontextu

Skrytá markovská posloupnost (také řada, řetěz) představuje statistický model (zkracováno HMM).

- Místo jednoho rozhodnutí jde o posloupnost rozhodnutí. Příští rozhodnutí závisí na předchozích rozhodnutích.
- Obvyklé aplikace: analýza pozorování měnících se v čase. Například: řečový signál, časová posloupnost tahů rukopisu.
- Skryté markovské modely (Hidden Markov Model – zkratka HMM) jsou "zlatým standardem" v analýze časových řad.
- Důvod?

  Skrytý markovský řetěz je statistický model, pro nějž ještě existuje algoritmus s polynomiální složitostí (algoritmus dynamického programování).

Skryté markovské pole zobecňuje statistický model pro mřížkové struktury (např. pixely v obrázku), tedy nejbližší trochu složitější. Zde už nejsou k dispozici algoritmy s polynomiální složitostí.

Značení posloupností zjednodušující výrazy

- Při popisu HMM budeme pracovat s posloupnostmi pozorování $\bar{x}$ a s posloupnostmi skrytých stavů $\bar{y}$.

$$\begin{array}{lllll} \text{pozorování} & \bar{x} = (& x_1,\ x_2,\ \ldots,\ x_n) & \in X^n \\ \text{skryté stavy} & \bar{y} = (y_0, & y_1,\ y_2,\ \ldots,\ k_n) & \in Y^{n+1} \end{array}$$

- Zaveďme stručnější značení posloupností: $\bar{x} = (x_a,\ x_{a+1},\ \ldots,\ x_b) = x_a^b$, které výrazy zjednoduší.

- Například výše zavedená posloupnost pozorování $\bar{x}$ a skrytých stavů $\bar{y}$ se zjednoduší na

$$\begin{array}{ll} \text{pozorování (stručně)} & \bar{x} = x_1^n \\ \text{skryté stavy (stručně)} & \bar{y} = y_0^n \end{array}$$

Příklady aplikačních oblastí

- Rozpoznávání řečových signálů ($\bar{x}$ signál z mikrofonu, $\bar{y}$ fonémy).
- Vyhledávání slov v promluvě ($\bar{x}$ slova, $\bar{y}$ vyhledávaný vzor, typicky slovo nebo několik slov).
- Rozpoznávání rukopisných znaků, ($\bar{x}$ on-line tahy pera, $\bar{y}$ např. jednotlivé znaky nebo vzorové podpisy).
- Biomedicínské inženýrství, analýza EKG a EEG signálů, ($\bar{x}$ signál, $\bar{y}$ charakteristiky signálu).
- Bioinformatika, analýza DNA sekvencí ($\bar{x}$ odezvy fluorescenčně označených molekul, $y \in \{A, C, G, T\}$) nebo ($x \in \{A, C, G, T\}$, $\bar{y}$ interpretačně významné podposloupnosti).
- Mobilní robotika ($\bar{x}$ body trajektorie robotu, $\bar{y}$ interpretace trajektorie).
- Rozpoznávání v obrazech, ale musí být zvláštní s jednorozměrným uspořádáním, např. rozpoznávání registračních značek aut ($\bar{x}$ sloupce obrazu registrační značky, $\bar{y}$ znaky a symboly na registrační značce).

Doporučené čtení

- Schlesinger M.I., Hlaváč V.: Deset přednášek z teorie statistického a strukturního rozpoznávání, monografie, Vydavatelství ČVUT, Praha 1999, 521 s.
- Rabiner L.R.: A tutorial on Hidden Markov Models and selected applications in speech recognition, časopis Proceedings of the IEEE, Vol. 77, No. 2, 1989, pp. 257-286.

Andrej Andrejevič Markov

- Narozen v Rjazani 1856, zemřel 1922 v Petrohradě.
- Ruský matematik, profesor Univerzity v Petrohradě, člen Ruské akademie věd, studoval stochastické procesy. Rozhodující článek z 1912.
- Markovské řetězy: posloupnosti náhodných proměnných, hodnota následující proměnné je určena hodnotou předchozí proměnné, ale nezávislá na předchozích stavech.
- Andrej Markov použil markovské řetězy pro studium distribuce metodu na analýzu distribuce samohlásek ve verších Evžen Oněgin básníka Sergeje Puškina.

Markovské modely a automaty

- Markovské modely (včetně skrytých, HMM) jsou zvláštním případem stochastických konečných automatů.
- Markovské modely dovolují vyjádřit statistické závislosti dané pořadím pozorování (stavů) jako například v časových řadách.

Konečný automat $(Y, V, X, \delta, k_0, F)$

- Y - konečná množina stavů automatu (skryté stavy);
- V - konečná abeceda vstupních symbolů;
- X - konečná abeceda výstupních symbolů;
- y_0 - počáteční (skrytý) stav, $y_0 \in Y$;
- F - cílové stavy; $F \subset Y$;
- δ - přechodová funkce stavů; $\delta\colon Y \times V \to Y \times X$.

- Když je automat ve stavu $y \in Y$ a na jeho vstup se přivede symbol $v \in V$, automat přejde do stavu $y' \in Y$ a na svém výstupu generuje výstupní symbol $x \in X$.
- Přechodová funkce δ určí hodnoty dvojice (y', x).
- Automat pracuje v krocích dokud nedosáhne stavu $y' \in F$, což je ukončovací podmínka.

Stochastický konečný automat

- Konečný automat zobecníme tím, že přechody ze stavu do stavu budou náhodné.
- Počáteční stav je také náhodný. Je dán apriorní prvděpodobností počátečního stavu $p(y_0)\colon Y \to \mathbb{R}$.
- Přechodová funkce stavů $\delta\colon Y \times V \to Y \times X$ se zobecní na stochastickou přechodovou funkci $\delta_s\colon Y \times V \to Y \times X$. Dolní index u δ_s označuje, že přechodová funkce stavu je stochastická.
- Znamená to, že odpovídající přechodová funkce stavu je náhodná. Nový stav a výstup bude $(y',\, x) = \delta_s(y, v)$, kde výstupní hodnoty jsou generovány v souladu s podmíněnou pravděpodobnostní distribucí $p(y',\, x \,|\, y, v)$.

Autonomní stochastický konečný automat

- V této přednášce se věnujeme zvláštnímu případ stochastického konečného automatu, kterému se říká autonomní stochastický konečný automat.
- Jeho vstupní abeceda V obsahuje jediný symbol, což vyjadřuje zvláštní případ, kdy chování automatu nezávisí na vstupních symbolech.
- Později uvidíme, že autonomní stochastický konečný automat je ekvivalentní se (skrytými) markovskými řetězy.
- Přechod mezi stavy je podobný jako u obecného stochastického konečného automatu. Jen je nezávislý na vstupních symbolech.
- Uvažujme množinu stavů Y a množinu výstupních symbolů X. Přechody mezi stavy jsou určeny pravděpodobnostmi $p(y_0)$, $p(x_i, y_i \,|\, y_{i-1})$, $y_0 \in Y$, $y_i \in Y$, $x_i \in X$, $i = 1, 2, \ldots, n$.

Fungování (autonomního) stochastického automatu

V zavedeném (autonomním) stochastickém automatu se přechod mezi skrytým stavem y_i a následujícím skrytým stavem y_{i+1} řídí rozdělením pravděpodobnosti

$$p(\bar{x}, \bar{y}) = p(x_1, \ldots, x_n, y_0, \ldots, y_n) = p(y_0) \prod_{i=1}^{n} p(x_i, y_i | y_{i-1})$$

To znamená, že automat:

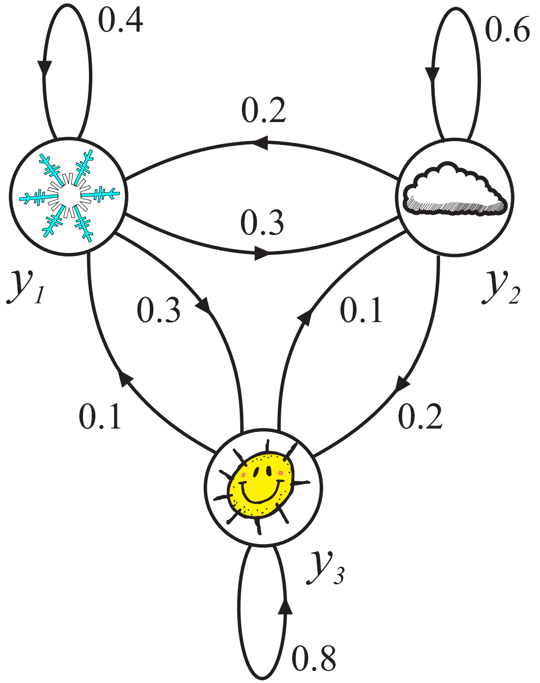
- na počátku generuje náhodný stav y_0 s pravděpodobností $p(y_0)$ a přejde do něj,
- v $i-$tém kroku generuje dvojici (x_i, y_i) s pravděpodobností $p(x_i, y_i | y_{i-1})$. Na výstupu dává symbol x_i a přechází do stavu y_i.

Příklad: generativní markovský model počasí (1)

Stochastický konečný automat předpovídající počasí aniž by něco měřil.

- Stav $q = y_1$: dešťové nebo sněhové srážky.
- Stav $q = y_2$: zataženo.
- Stav $q = y_3$: slunečno.
- Přechodová matice stavů A mezi stavem q_t a předchozím stavem q_{t-1} je nezávislá na čase.

  $a_{ij} = p(q_t = y_i | q_{t-1} = y_j),\ i, j \in 1, 2, 3$



$$p(q_t = y_i) = A\, p(q_{t-1} = y_{i-1}) = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} p(q_{t-1} = y_{i-1})$$

Poděkování: Ricardo Gutierrez-Osuna

Příklad: generativní markovský model počasí (2)

- *Otázka 1*: Je zadáno, že ve dni $t = 1$ slunečno. Jaká je pravděpodobnost že posloupnost počasí v následujících sedmi dnech bude "slunečno, slunečno, déšť, déšť, slunečno, zataženo, slunečno"?

  Odpověď 1:

$$
\begin{aligned}
& p(y_3, y_3, y_3, y_1, y_1, y_3, y_2, y_3 | \text{statistický model}) = \\
& \quad = p(y_3)\, p(y_3|y_3)\, p(y_3|y_3)\, p(y_1|y_3)\, p(y_1|y_1)\, p(y_3|y_1)\, p(y_2|y_3)\, p(y_3|y_2) = \\
& \quad = p(y_3)\, a_{33}\, a_{33}\, a_{13}\, a_{11}\, a_{31}\, a_{23}\, a_{32} = \\
& \quad = 1 \cdot 0.8 \cdot 0.8 \cdot 0.1 \cdot 0.4 \cdot 0.3 \cdot 0.1 \cdot 0.2 = 0.0001536
\end{aligned}
$$

- *Otázka 2*: Jaká je pravděpodobnost, že počasí, stav y_i, zůstane stejné po přesně m po sobě následujících dní?

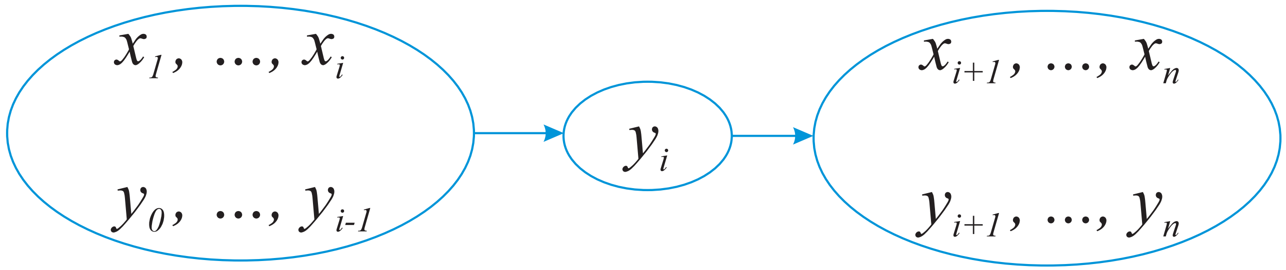
  Odpověď 2: $p(q_t = y_i, q_{t+1} = y_i \dots q_{t+m} = y_{j \neq i}) = a_{ii}^{m-1}(1 - a_{ii})$

Markovské posloupnosti se skrytými stavy

- Statistický model $p(\bar{x}, \bar{y}) = X^n \times Y^{n+1} \to \mathbb{R}$.
- Markovský řetěz:

  Předpokládáme, že pro všechny posloupnosti $\bar{x} = (x_1^i, x_{i+1}^n)$ a $\bar{y} = (y_0^{i-1}, y_i, y_{i+1}^n)$ platí

$$p(\bar{x}, \bar{y}) = p(y_i)\, p(x_1^i, y_0^{i-1} | y_i)\, p(x_{i+1}^n, y_{i+1}^n | k_i)\ . \qquad (1)$$



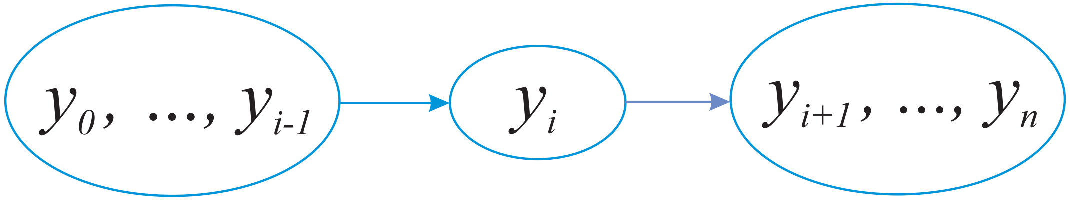
Markovská podmínka pro (skryté) stavy

- Vyjdeme z markovské podmínky

$$p(\bar{x}, \bar{y}) = p(y_i)\, p(x_1^i, y_0^{i-1} | y_i)\, p(x_{i+1}^n, y_{i+1}^n | y_i)\,.$$

- Pro skryté stavy po sčítání (marginalizaci) přes všechna možná pozorování x platí markovská vlastnost posloupnosti skrytých stavů $\bar{y}$

$$p(\bar{y}) = p(y_i)\, p(y_0^{i-1} | y_i)\, p(y_{i+1}^n | y_i)\,.$$



Skryté markovské posloupnosti (2)

V rovnici $p(\bar{x},\bar{y}) = p(y_i)\, p(x_1^i, y_0^{i-1}|y_i)\, p(x_{i+1}^n, k_{i+1}^n|y_i)$ budeme sčítat přes posloupnost skrytých stavů y_{i+2}^n a potom přes posloupnost pozorování x_{i+2}^n a získáme

$$\begin{aligned} p(x_1^{i+1}, y_0^{i+1}) &= \sum_{x_{i+2}^n} \sum_{y_{i+2}^n} p(x,y) = p(x_1^i, y_0^i) \sum_{x_{i+2}^n} \sum_{y_{i+2}^n} p(x_{i+1}^n, y_{i+1}^n|y_i) \\ &= p(x_1^i, y_0^i)\, p(x_{i+1}, y_{i+1}|y_i) \end{aligned}$$

Využijeme předchozí vztah rekurzivně a získáme

$$p(\bar{x},\bar{y}) = p(x_1,\ldots,x_n,y_0,\ldots,y_n) = p(y_0) \prod_{i=1}^{n} p(x_i, y_i|y_{i-1})$$

Výpočet složité funkce $2\,n+1$ proměnných se zjednodušil na výpočet n funkcí $p(x_i, y_i|y_{i-1})$ o třech proměnných a jedné funkce $p(k_0)$ o jedné proměnné.

Interpretace markovské vlastnosti

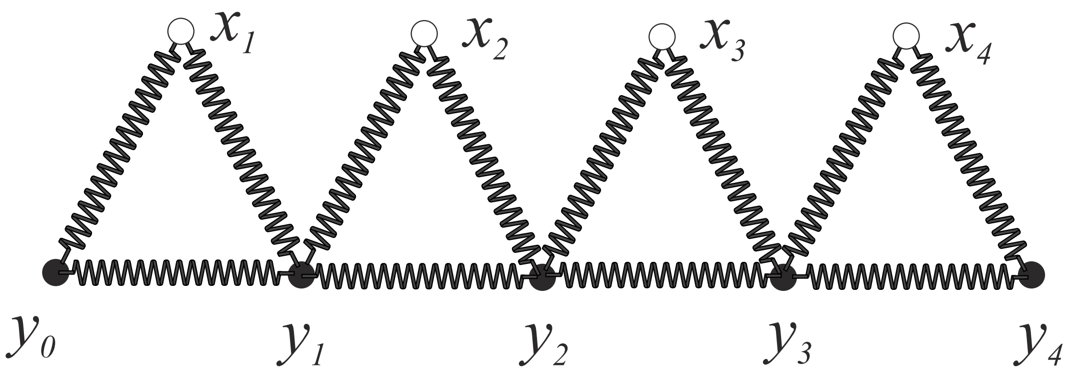
- Uvažujme všechny možné dvojice posloupností (x_1^n, y_0^n), které splňují markovskou podmínku (1) z průsvitky 14.
- Zvolme libovolnou hodnotu i, $0 < i < n$. Zvolme libovolnou hodnotu skrytého parametru $y_i = \sigma$.
- Z možných dvojic markovských posloupností (x_1^n, y_0^n) vyberme soubor posloupností, pro něž platí $y_i = \sigma$.
- Být markovkou posloupností potom znamená, že parametry $(x_1, x_2, \ldots, x_i)$, $(y_0, y_1, \ldots, y_{i-1})$ ve vybraném souboru posloupností jsou statisticky nezávislé na parametrech $(x_{i+1}, x_{i+2}, \ldots, x_n)$, $(y_{i+1}, y_{i+2}, \ldots, y_n)$.

Pozor na nesprávnou interpretaci

- Často se uvádí nepřesně zjednodušená interpretace: “Markovská posloupnost je taková, u níž nezáleží na minulosti, ale jen na současnosti.”
- Tato interpretace je zrádná, protože je sice nesprávná, ale od správné se liší jen málo.
- Ilustrujme na příští průsvitce tuto situaci na mechanickém modelu markovské posloupnosti.

Mechanický model markovské posloupnosti

- Uvažujme posloupnosti x_1^4 a y_0^4 reprezentované vrcholy rovinného grafu. Některé vrcholy grafu jsou spojeny pružinami označujícími statistickou vazbu v markovském modelu.



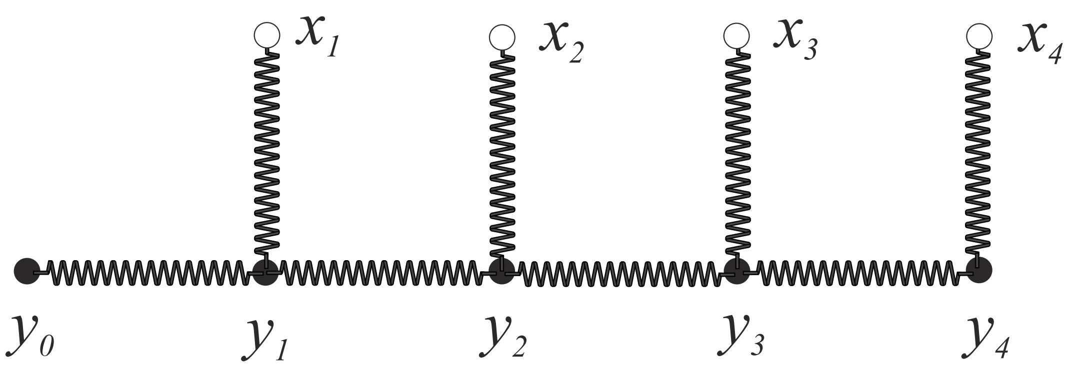
- Představme si, že např. vrchol grafu x_3 začne oscilovat. Díky vazbám postupně začnou oscilovat všechny ostatní body.
- Pokud jsou hodnoty $x_1, \ldots, x_4$ fixovány, jsou určeny i hodnoty $y_0, \ldots, y_n$.
- Když zafixujme např. vrchol y_3, potom se model rozloží na dvě nezávislé části: (a) vrcholy y_0, y_1, y_2, x_1, x_2, x_3. (b) vrcholy x_4, y_4.

Zvláštní případ: Dekomponovatelný statistický model

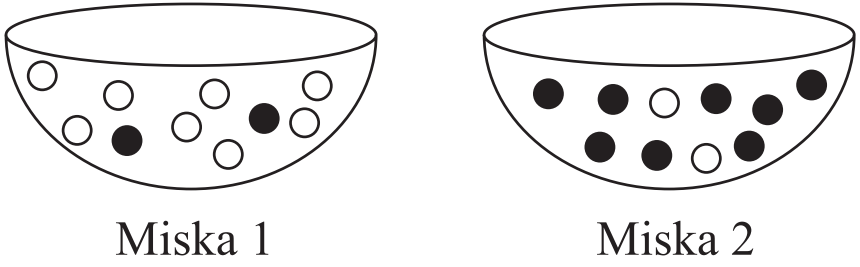
- Jde o zvláštní příklad často uvažovaný v literatuře.
- Předpokládá se: $p(x_i, k_i|k_{i-1}) = p(x_i|k_i)\, p(k_i|k_{i-1})$. Potom platí

$$p(\bar{x}, \bar{y}) = p(y_0) \prod_{i=1}^{n} p(x_i, y_i|y_{i-1}) = p(y_0) \prod_{i=1}^{n} p(x_i|y_i) \prod_{i=1}^{n} p(y_i|y_{i-1})$$

- Odpovídající mechanický model



Příklad dekomponovatelného modelu
Tahání kuliček z misek



kulička $x = \{$černá, bílá$\}$
miska $y = \{1, 2\}$

$p(y=1) = 0.5$ $\quad p(x=\text{bílá}|y=1) = 0.8 \quad p(x=\text{bílá}|y=2) = 0.2$

$p(y=2) = 0.5$ $\quad p(x=\text{černá}|y=1) = 0.2 \quad p(x=\text{černá}|y=2) = 0.8$

Tahání z misek střídavě

$p(y=1|y=2) = 1 \quad p(y=2|y=2) = 0$

$p(y=1|y=1) = 0 \quad p(y=2|y=1) = 1$

Tři základní úlohy s HMM

1. Rozpoznávání též ohodnocení statistického modelu:

   (v angl. literatuře algoritmus forward-backward, dynamické programování). Dány parametry HMM (statistických modelů), cílem je spočítat pravděpodobnost, že je pozorována posloupnost $\bar{x}$. Třída odpovídá nejpravděpodobnějšímu modelu. Používá se při rozpoznávání.

2. Hledání nejpravděpodobnější posloupnosti skrytých stavů:

   (Viterbi algoritmus, dynamické programování). Dán statistický model a posloupnost pozorování $\bar{x}$. Cílem je najít nejpravděpodobnější posloupnost skrytých stavů $\bar{y}$.

3. Učení statistického modelu:

   (Baum-Welsh algoritmus, využívá EM algoritmus). Dána struktura modelu, tj. počet skrytých stavů a trénovací množina. Cílem je najít parametry statistického modelu, tj. pravděpodobnosti $p(x_i, y_i | y_{i-1})$.

Úloha rozpoznávání; též ohodnocení určitého statistického modelu

- Nechť a, b jsou dva autonomní stochastické automaty se stejným počtem stavů Y a výstupních symbolů X. Počet výstupních symbolů $|X|$ je pro oba automaty také totožný.
- Nechť se statistické vlastnosti automatů a, b liší.

  Automat a je popsán: $p_a(y_0)$ a $p_a(x_i, y_i \mid y_{i-1})$, $y_0 \in Y$, $y_i \in Y$, $x_i \in X$, $i = 1, 2, \ldots, n$.

  Podobně automat b je popsán: $p_b(y_0)$ a $p_b(x_i, y_i \mid y_{i-1})$.

  Poznámka: Zde pro jednoduchost pravděpodobnosti nezávisí na indexu i. V obecném případě mohou pravděpodobnosti na indexu i záviset. Naše úvahy platí také pro obecnější případ.
- Úloha ohodnocení statistického modelu (též úloha rozpoznávání) má při znalosti statistických modelů automatů rozhodnout, který automat generoval posloupnost pozorování $x_1, x_2, \ldots, x_n$.

Úloha rozpoznávání (2)

- Úlohu rozpoznávání lze vyjádřit jako minimalizaci bayesovského rizika (např. pro jednoduchost pravděpodobnost chybného rozhodnutí). Formulace může být rozšířena na nebayesovské úlohy, např. Neyman-Pearsonovu úloha, minimaxní úlohu.
- Je potřebné spočítat odpovídající marginální pravděpodobnosti pro automaty a, b $p_a(x_1^n)$ a $p_b(x_1^n)$.
- Rozhodne se např. podle maximální věrohodnosti $p_a(x_1^n)\ /\ p_b(x_1^n)$.

  Vzpomeňme si na přednášku o nebayesovském rozpoznávání, úloha Neyman-Pearson pro dichotomickou klasifikaci.
- Nejnáročnější částí úlohy je spočítat pravděpodobnosti $p_a(x_1^n)$ a $p_b(x_1^n)$. Výpočet je stejný pro automaty a i b, a tak neuvádíme index určující konkrétní automat.

Algoritmus pro úlohu rozpoznávání

Vzpomeňme na markovský statistický model

$$p(\bar{x},\bar{y}) = p(x_1,\ldots,x_n,y_0,\ldots,y_n) = p(y_0)\prod_{i=1}^{n} p(x_i,y_i|y_{i-1})$$

Nás nyní zajímá marginální pravděpodobnost $p(\bar{x}) = \sum_{y\in Y} p(\bar{x},\bar{y})$. Vyjádří se jako mnohočetná suma

$$p(x) = \sum_{y} p(\bar{y},\bar{x}) = \sum_{y_0}\sum_{y_1}\cdots\sum_{y_{n-1}}\sum_{y_n} p(y_0)\prod_{i=1}^{n} p(x_i,y_i\,|\,y_{i-1})\,.$$

Přímý výpočet není praktický, protože sčítanců je $|Y|^{n+1}$. Vzorec lze ekvivalentními transformacemi upravit do použitelnějšího tvaru.

Algoritmus rozpoznávání (2)

Při sčítání lze podle stavu y_i vytknout činitele nezávisející na y_i. Vyjdeme z již známého

$$p(\bar{x}) = \sum_k p(\bar{y},\bar{x}) = \sum_{y_0} \sum_{y_1} \cdots \sum_{y_{n-1}} \sum_{y_n} p(y_0) \prod_{i=1}^{n} p(x_i, y_i \,|\, y_{i-1}) \,.$$

Po vytýkání se převede na

$$\begin{aligned} p(\bar{x}) \quad &= \quad \sum_{y_0} p(y_0) \sum_{y_1} p(x_1, y_1 \,|\, y_0) \cdots \sum_{y_i} p(x_i, y_i \,|\, y_{i-1}) \\ &\cdots \quad \sum_{y_{n-1}} p(x_{n-1}, y_{n-1} \,|\, y_{n-2}) \sum_{k_n} p(x_n, y_n \,|\, y_{n-1}) \,. \end{aligned}$$

Algoritmus rozpoznávání (3)

Míříme k rekurzivnímu algoritmu. Ve vztahu po vytýkání

$$
\begin{aligned}
p(\bar{x}) &= \sum_{y_0} p(y_0) \sum_{y_1} p(x_1, y_1 \,|\, y_0) \cdots \sum_{y_i} p(x_i, y_i \,|\, y_{i-1}) \\
&\cdots \sum_{y_{n-1}} p(x_{n-1}, y_{n-1} \,|\, y_{n-2}) \sum_{y_n} p(x_n, y_n \,|\, y_{n-1}) \,.
\end{aligned}
$$

si označíme dílčí součty pro $i = 1, 2, \ldots, n$ pomocí

$$
\begin{aligned}
f_i(y_{i-1}) &= \sum_{y_i} p(x_i, y_i \,|\, y_{i-1}) \sum_{y_{i+1}} p(x_{i+1}, y_{i+1} \,|\, y_i) \cdots \\
&\cdots \sum_{y_{n-1}} p(x_{n-1}, y_{n-1} \,|\, y_{n-2}) \sum_{k_n} p(x_n, y_n \,|\, y_{n-1})
\end{aligned}
$$

a získáme algoritmus výpočtu.

Algoritmus rozpoznávání (4)

Výpočet probíhá odzadu posloupnosti

$$\left.\begin{aligned} f_n(y_{n-1}) &= \sum_{y_n} p(x_n, y_n \,|\, y_{n-1}); \\ f_i(y_{i-1}) &= \sum_{y_i} p(x_i, y_i \,|\, y_{i-1})\; f_{i+1}(y_i)\;, \quad i = 1, 2, \ldots, n-1\,; \\ p(\bar{x}) &= \sum_{y_0} p(y_0)\; f_1(y_0)\;. \end{aligned}\right\}$$

Počet operací při výpočtu je úměrný $|Y|^2\, n$.

Nejpravděpodobnější posloupnost skrytých stavů

Formulace úlohy

- Je dán statistický model $p(\bar{x}, \bar{y}) = p(y_0) \prod_{i=1}^{n} p(x_i, y_i | y_{i-1})$.
- Hledá se rozhodovací strategie $q(\bar{x})\colon X^n \to Y^{n+1}$.
- Bayesovské riziko je $R(q(\bar{x})) = \sum_{\bar{x} \in X} \sum_{\bar{y} \in Y} p(\bar{x}, \bar{y})\, W(\bar{x}, q(\bar{x}))$.
- Jednoduchá pokutová funkce, tj. počet chybných rozhodnutí

$$W(\bar{y}, q(\bar{x})) = \begin{cases} 0 & \text{pro } \bar{y} = q(\bar{x}) \,, \\ 1 & \text{pro } \bar{y} \neq q(\bar{x}) \,. \end{cases}$$

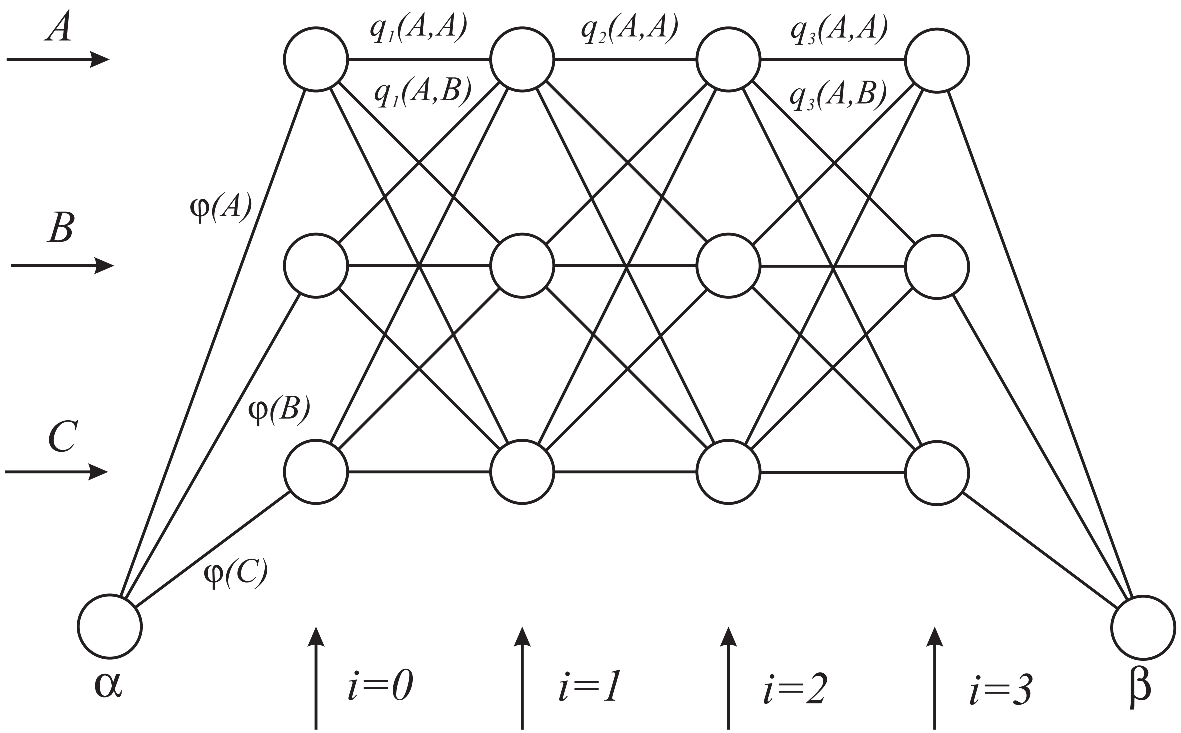
Cílem je najít strategii $q(\bar{x})$ minimalizující riziko $R(q(\bar{x}))$.

Odvození bayesovské strategie

$$
\begin{aligned}
q(\bar{x}) &= \operatorname*{argmax}_{y \in Y^{n+1}} \frac{p(\bar{x},\bar{y})}{\sum\limits_{y' \in Y^{n+1}} p(\bar{x},\bar{y}')} = \operatorname*{argmax}_{y \in Y^{n+1}} \; p(\bar{x},\bar{y}) \\
&= \arg\max_{y_0} \ldots \max_{y_n} \; p(y_0) \prod_{i=1}^{n} p(x_i, y_i | y_{i-1}) \\
&= \arg\max_{y_0} \ldots \max_{y_n} \; \log \left(p(y_0) \prod_{i=1}^{n} p(x_i, y_i | y_{i-1}) \right) \\
&= \arg\max_{y_0} \ldots \max_{y_n} \; \left(\underbrace{\log p(y_0)}_{\varphi(y_0)} + \sum_{i=1}^{n} \underbrace{\log p(x_i, y_i | y_{i-1})}_{q_i(y_{i-1}, y_i)} \right)
\end{aligned}
$$

Formulace přes hledání nejkratší cesty v grafu

- Orientovaný graf s vrcholy a hranami orientovanými zleva doprava mezi nimi. Počátečním vrcholem je α a cílovým vrcholem β. Zbylých $|Y|(n+1)$ mezilehlých vrcholů indexujeme dvojicí (σ, i), $\sigma \in Y$, $i = 0, 1, \ldots, n$.
- Příklad grafu pro $n = 3$ a množinu skrytých stavů $Y = \{A, B, C\}$.



Algoritmus hledání nejkratší cesty

Dynamické programování. Graf má zvláštní tvar. Je zaručeno uspořádání. (analogie – poslové).

$f_i(\sigma)$ je délka nejkratší cesty ($\sim$ času) z vrcholu α do (σ, i).

Algoritmus A. Viterbi 1967 (nezávisle T. Vincjuk 1968):

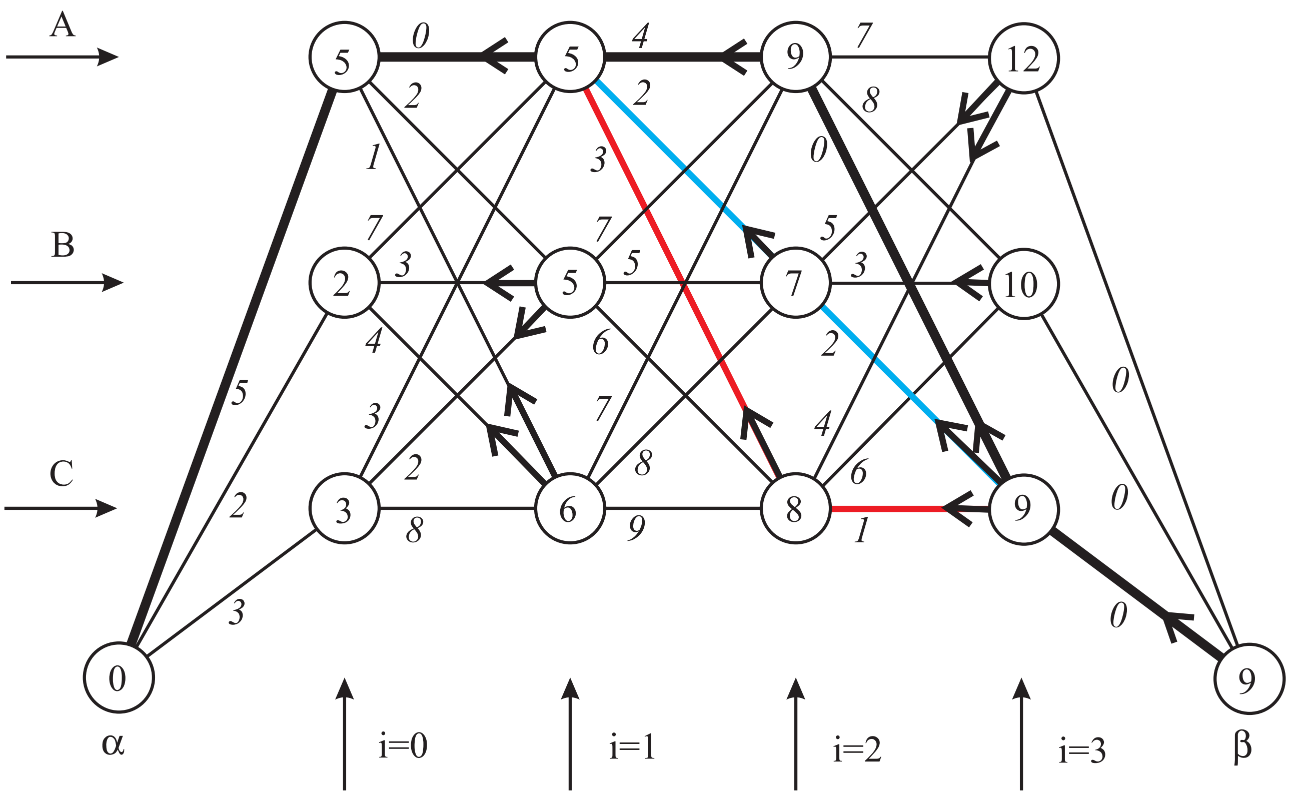
- $f_0(\sigma) = \varphi(\sigma)$
- Opakovaně pro $i = 1, \ldots, n, \quad \sigma \in Y$

  $f_i(\sigma) = \min\limits_{\sigma' \in Y} \left(f_i(\sigma') + q_i(\sigma', \sigma)\right)$. Dráha posla, který přišel nejdříve.

  $\mathrm{ind}_i(\sigma) = \operatorname*{argmin}\limits_{\sigma' \in y} \left(f_i(\sigma') + q_i(\sigma', \sigma)\right)$. Vrchol, z něhož přišel první posel.
- Nakonec:

  $y_n = \operatorname*{argmin}\limits_{\sigma \in Y} f_n(\sigma), \quad y_{i-1} = \mathrm{ind}_i(y_i)$. Rekonstrukce nejkratší cesty.

Příklad na Viterbiho algoritmus



Šipky označují $\mathrm{ind}_i(\sigma)$. Šipky se použijí pro nalezení optimální cesty. Těchto cest může být více. V našem příkladě jsou to kromě **AAAC** i AABC nebo AACC.