Implementace Bayesova kasifikátoru a diskriminačních funkcí v prostředí Matlab

J. Havlík

Katedra teorie obvodů
Fakulta elektrotechnická
České vysoké učení technické v Praze
Technická 2, 166 27 Praha 6

email: xhavlikj@fel.cvut.cz

Abstrakt

Příspěvek se zabývá problematikou klasifikace prostorových pohybů palce ruky. Palec ruky je označen reflexní značkou, jejíž pohyb je sledován dvojicí videokamer ze dvou lineárně nezávislých směrů. Pohyb značky v prostoru je parametrizován a uložen jako posloupnost polohových vektorů. V příspěvku je uveden popis Bayesova klasifikátoru a klasifikátoru založeného na diskriminačních funkcích a tyto dva přístupy jsou porovnány.

1 Úvod

Uvedené klasifikační algoritmy jsou částí řetězce použitého k výzkumu vzájemných souvislostí svalové a mozkové aktivity člověka.

Mozková aktivita je v tomto výzkumu reprezentována elektro-encefalografickým signálem (EEG). Signál EEG neobsahuje pouze části přímo související se sledovaným pohybem těla, ale obsahuje i části odrážející další životní funkce, např. artefakty očních pohybů. Aby bylo možné provádět srovnání EEG záznamu se záznamem svalové aktivity, bylo třeba zvolit jednoduchý pohyb malého svalu, který při pohybu příliš neovlivňuje své okolí. Jako sledovaný pohyb byl vybrán volný pohyb palce ruky. Svalová aktivita je reprezentována posloupností polohových vektorů značky umístěné na palci ruky.

Účelem celého řetězce je nalézt typické změny mozkové aktivity odpovídající specifickým pohybům palce ruky.

2 Experiment

Snímaná osoba hýbe prstem mezi 3 polohami, směr pohybu prstu je volen pokusnou osobou tak, aby všechny směry pohybu byly realizovány zhruba se stejnou četností. Pohyb prstu následuje vždy po optickém synchronizačním pulsu, perioda pulsů je 6 ± 1 vteřina.

Snímaný pohyb je zaznamenáván dvojicí standardních DV videokamer (rozlišení 720×520 pixelů, snímkovací frekvence 25 snímků za vteřinu) ze dvou lineárně nezávislých směrů. Uspořádání měřicího pracoviště je zřejmé z obrázku 1.

Palec ruky je označen speciální reflexní značkou. Pohyb této značky v zaznamenaných videosekvencích je parametrizován tak, aby bylo možné v každém půlsnímku zvlášť určit polohu této značky. Z každé vteřiny záznamu je tak získána posloupnost 50 polohových vektorů reprezentujících polohu palce ruky v prostoru. Tyto vektory, označme je $\boldsymbol{x}$, jsou dále použity jako příznakové vektory pro klasifikaci pohybů prstu.

3 Bayesův klasifikátor

3.1 Klasifikační algoritmus

Řešíme klasifikaci vstupních vektorů $\boldsymbol{x}$ do R tříd, indikátory tříd označme $\omega_1, \omega_2, \ldots, \omega_R$.

Apriorní pravděpodobnosti $P(\omega_1), P(\omega_2), \ldots, P(\omega_R)$ jsou známy a platí pro ně

$$\sum_{r=1}^{R} P(\omega_r) = 1. \tag{1}$$

Apriorní pravděpodobnosti udávají, jaká je pravděpodobnost, že vektor $\boldsymbol{x}$ připadne do třídy s indikátorem ω_r.

Předpokládejme, že kromě apriorních pravděpodobností $P(\omega_r)$ známe ještě všechny podmíněné hustoty pravděpodobnosti $p(\boldsymbol{x}|\omega_r)$ udávající rozložení hodnot $\boldsymbol{x}$ uvnitř jednotlivých tříd. Pak platí, že vektor příznaků $\boldsymbol{x}$ patří do třídy s indikátorem ω_r s pravděpodobností

$$P(\omega_r|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\omega_r)P(\omega_r)}{p(\boldsymbol{x})}, \tag{2}$$

kde

$$p(\boldsymbol{x}) = \sum_{r=1}^{R} p(\boldsymbol{x}|\omega_r)P(\omega_r). \tag{3}$$

Pravděpodobnosti $P(\omega_r|\boldsymbol{x})$ nazýváme *aposteriorními pravděpodobnostmi*, vztah (2) nazýváme *Bayesovým vztahem* [3].

Obrázek 1: Měřicí pracoviště

Vektor příznaků $\boldsymbol{x}$ zařadíme do té třídy ω^*, pro níž platí

$$\omega^* = \arg\max_r P(\omega_r|\boldsymbol{x}). \tag{4}$$

Vzhledem k tomu, že jmenovatel Bayesova vztahu (2) je pro všechny aposteriorní pravděpodobnosti (třídy) stejný, je možné provádět klasifikaci na základě velikosti čitatele tohoto zlomku.

Jak vyplývá z definice experimentu (pokusné osoby jsou žádány, aby stacionární stavy mezi pohyby prstu náhodně střídaly tak, aby relativní četnost jednotlivých tříd pohybů byla zhruba stejná), apriorní pravděpodobnosti všech tříd jsou si rovny a platí pro ně

$$P(\omega_r) = \frac{1}{R} \quad \text{pro všechna} \quad r = 1, \ldots, R. \tag{5}$$

Z uvedeného (5) vyplývá, že i čitatel Bayesova vztahu (2) je pro účely klasifikace možné dále zjednodušit a vlastní klasifikaci provádět pouze na základě velikostí podmíněných hustot pravděpodobností

$$\omega^* = \arg\max_r p(\boldsymbol{x}|\omega_r). \tag{6}$$

3.2 Trénování klasifikátoru

Ke správné klasifikaci příznakových vektorů $\boldsymbol{x}$ do tříd $\omega_1, \omega_2, \ldots, \omega_R$ potřebujeme, jak vyplývá z (6), znát podmíněné hustoty pravděpodobností $p(\boldsymbol{x}|\omega_r)$.

Předpokládejme, že hustoty pravděpodobností $p(\boldsymbol{x}|\omega_r)$ udávající rozložení hodnot příznakových vektorů $\boldsymbol{x}$ v jednotlivých třídách odpovídají normálnímu rozložení a je tedy možné je popsat rovnicí

$$p(\boldsymbol{x}|\omega_r) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}_r|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_r)' \boldsymbol{\Sigma}_r^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_r)\right), \tag{7}$$

kde

N	je počet prvků příznakového vektoru $\boldsymbol{x}$,
$\boldsymbol{\mu}_r$	je střední hodnota normálního rozdělení,
$\boldsymbol{\Sigma}_r$	je jeho kovarianční matice. [4]

Pro získání hustot pravděpodobností $p(\boldsymbol{x}|\omega_r)$ pak potřebujeme znát pro každou klasifikovanou třídu hodnoty parametrů $\boldsymbol{\mu}_r$ a $\boldsymbol{\Sigma}_r$. Tyto parametry získáme na základě statistického rozboru příznakových vektorů $\boldsymbol{x}$, u nichž známe, ke které klasifikační třídě patří.

3.3 Odhad parametrů normálního rozdělení

Předpokládejme, že máme trénovací množinu obsahující M vzájemně statisticky nezávislých příznakových vektorů $\boldsymbol{x}$ patřících k jedné třídě s identifikátorem ω_r. Pak pro danou trénovací množinu můžeme psát

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M | \boldsymbol{q}) = \prod_{m=1}^{M} p(\boldsymbol{x}_m | \boldsymbol{q}). \tag{8}$$

Hustota pravděpodobnosti na levé straně rovnice (8) je pro danou trénovací množinu vektorů $\boldsymbol{x}$ závislá pouze na parametru $\boldsymbol{q}$ a nazýváme ji *funkcí věrohodnosti*. Úkolem trénování klasifikátoru je nalézt pro danou třídu ω_r takový parametr $\boldsymbol{q}^*$, který maximalizuje věrohodnost (8), tedy

$$\boldsymbol{q}^* = \arg\max_{\boldsymbol{q}} \prod_{m=1}^{M} p(\boldsymbol{x}_m|\boldsymbol{q}). \tag{9}$$

Takovýto odhad parametru $\boldsymbol{q}$ nazýváme *maximálně věrohodný odhad* (MLE - Maximum Likelihood Estimation).

Dosazením předpokládaného tvaru hustoty pravděpodobnosti $p(\boldsymbol{x}|\omega)$ (7) do rovnice pro věrohodnost (8) získáme podmíněnou hustotu pravděpodobnosti realizace trénovací množiny

$$\begin{aligned} p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \\ &= \prod_{m=1}^{M} \frac{1}{\sqrt{(2\pi)^N \, |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2} \left(\boldsymbol{x}_m - \boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{x}_m - \boldsymbol{\mu}\right) \right). \end{aligned} \tag{10}$$

Aniž bychom změnili polohu maxima (8), můžeme celou uvedenou rovnici zlogaritmovat (logaritmus je monotónně rostoucí funkce a nabývá tedy maxima tam, kde jeho argument)

$$\begin{aligned} \ln p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \\ &= -\frac{N \cdot M}{2} \ln(2\pi) - \frac{M}{2} \ln|\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{m=1}^{M} \left(\left(\boldsymbol{x}_m - \boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{x}_m - \boldsymbol{\mu}\right) \right). \end{aligned} \tag{11}$$

Maximalizaci věrohodnostní funkce provedeme nalezením extrému funkce v závislosti na parametrech $\boldsymbol{\mu}$ a $\boldsymbol{\Sigma}$.

Nalezení parametru $\boldsymbol{\mu}$

Provedeme parciální derivaci $\ln p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ podle vektoru $\boldsymbol{\mu}$

$$\begin{aligned} &\frac{\partial \ln p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} \\ &\quad = -\frac{1}{2} \sum_{m=1}^{M} \frac{\partial \left(\left(\boldsymbol{x}_m - \boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{x}_m - \boldsymbol{\mu}\right) \right)}{\partial \boldsymbol{\mu}} = \boldsymbol{\Sigma}^{-1} \sum_{m=1}^{M} \left(\boldsymbol{x}_m - \boldsymbol{\mu}\right). \end{aligned} \tag{12}$$

Výraz (12) položíme roven nule

$$\boldsymbol{\Sigma}^{-1} \sum_{m=1}^{M} \left(\boldsymbol{x}_m - \boldsymbol{\mu}\right) = 0 \tag{13}$$

a dostáváme

$$\boldsymbol{\mu} = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{x}_m. \tag{14}$$

Výraz (14) je maximálně věrohodným odhadem parametru $\boldsymbol{\mu}$ normálního rozdělení (7).

Nalezení parametru $\mathbf{\Sigma}$

U parametru $\mathbf{\Sigma}$ postupujeme obdobně jako u předchozího parametru. Nejprve vypočteme parciální derivaci věrohodnostní funkce podle tohoto parametru a položíme ji rovnou nule.

$$\begin{aligned}
&\frac{\partial \ln p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M \,|\, \boldsymbol{\mu}, \mathbf{\Sigma})}{\partial \mathbf{\Sigma}} = \\
&= -\frac{M}{2}\frac{\partial \ln |\mathbf{\Sigma}|}{\partial \mathbf{\Sigma}} - \frac{1}{2}\sum_{m=1}^{M} \frac{\partial \left((\boldsymbol{x}_m - \boldsymbol{\mu})' \, \mathbf{\Sigma}^{-1} \, (\boldsymbol{x}_m - \boldsymbol{\mu}) \right)}{\partial \mathbf{\Sigma}} = \\
&= -\frac{M}{2}\mathbf{\Sigma}^{-1} + \frac{1}{2}\mathbf{\Sigma}^{-1} \sum_{m=1}^{M} \left((\boldsymbol{x}_m - \boldsymbol{\mu}) \, (\boldsymbol{x}_m - \boldsymbol{\mu})' \right) \mathbf{\Sigma}^{-1} = 0
\end{aligned} \tag{15}$$

Poté řešením rovnice (15) vyjádříme parametr $\mathbf{\Sigma}$, pro nějž je hodnota věrohodnostní funkce (8) maximální.

$$\mathbf{\Sigma} = \frac{1}{M}\sum_{m=1}^{M} \left((\boldsymbol{x}_m - \boldsymbol{\mu}) \, (\boldsymbol{x}_m - \boldsymbol{\mu})' \right) \tag{16}$$

Kovarianční matice $\mathbf{\Sigma}$ je symetrická pozitivně definitní čtvercová matice s reálnými prvky σ_{ij}, kde $i, j = 1, \ldots, N$ [4].

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \ldots & \sigma_N^2 \end{pmatrix} \tag{17}$$

Prvky σ_i^2 na hlavní diagonále matice jsou rozptyly i-tého prvku vektorů $\boldsymbol{x}$, prvky σ_{ij} mimo hlavní diagonálu jsou kovariance i-tého a j-tého prvku těchto vektorů. Pokud jsou prvky příznakových vektorů $\boldsymbol{x}$ vzájemně nekorelované, je kovarianční matice $\mathbf{\Sigma}$ diagonální.

4 Diskriminační funkce

4.1 Klasifikační algoritmus

Řešíme opět klasifikaci vstupních vektorů $\boldsymbol{x}$ do R tříd, indikátory tříd označme $\omega_1, \omega_2, \ldots, \omega_R$.

Hledáme množinu funkcí $g_r(\boldsymbol{x})$, pro které platí

$$g_r(\boldsymbol{x}) > g_s(\boldsymbol{x}) \tag{18}$$

pro každý vektor $\boldsymbol{x}$ a pro $s = 1, 2, \ldots R, s \neq r$. Funkce splňující uvedenou nerovnost nazýváme *diskriminační funkce*, jejich výběr obvykle není jednoznačný [3].

Vektor příznaků $\boldsymbol{x}$ zařadíme do té třídy ω^*, pro níž platí

$$\omega^* = \arg\max_r g_r(\boldsymbol{x}). \tag{19}$$

S ohledem na rovnici (6) můžeme zvolit např. množinu funkcí

$$g_r(\boldsymbol{x}) = p(\boldsymbol{x}|\omega_r). \tag{20}$$

Je zřejmé, že platí-li

$$p(\boldsymbol{x}|\omega_r) > p(\boldsymbol{x}|\omega_s) \tag{21}$$

pro každý vektor $\boldsymbol{x}$ a pro $s = 1, 2, \ldots R, s \neq r$, pak za stejných podmínek platí i

$$\ln p(\boldsymbol{x}|\omega_r) > \ln p(\boldsymbol{x}|\omega_s). \tag{22}$$

Diskriminační funkce pak definujeme

$$g_r(\boldsymbol{x}) = \ln p(\boldsymbol{x}|\omega_r) \tag{23}$$

a dosadíme-li z rovnice (7), dostáváme

$$g_r(\boldsymbol{x}) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_r| - \frac{1}{2}\left(\boldsymbol{x}_m - \boldsymbol{\mu}_r\right)' \boldsymbol{\Sigma}_r^{-1}\left(\boldsymbol{x}_m - \boldsymbol{\mu}_r\right). \tag{24}$$

Po zanedbání konstant neovlivňujících nerovnost (18) můžeme předchozí rovnici přepsat do tvaru

$$g_r(\boldsymbol{x}) = -\ln|\boldsymbol{\Sigma}_r| - \left(\boldsymbol{x}_m - \boldsymbol{\mu}_r\right)' \boldsymbol{\Sigma}_r^{-1}\left(\boldsymbol{x}_m - \boldsymbol{\mu}_r\right). \tag{25}$$

Před vlastní klasifikací je třeba diskriminační klasifikátor rozhodující na základě vztahu (19) natrénovat, tedy pro jednotlivé třídy s identifikátory ω_r najít hodnoty parametrů $\boldsymbol{\mu}_r$ a $\boldsymbol{\Sigma}_r$. Při trénování postupujeme stejně jako při trénování Bayesova klasifikátoru.

4.2 Vlastnosti diskriminačních funkcí

Z odvození rovnice (23), zejména ze vztahů (20) a (22) je zřejmé, že použití takto definovaných diskriminačních funkcí vede k výsledkově totožnému klasifikačnímu algoritmu, jako je Bayesův klasifikátor popsaný v části 3.

Výhodou diskriminačních funkcí je jejich menší výpočetní náročnost ve srovnání s Bayesovým klasifikátorem, neboť ve vztahu (25) není třeba vypočítávat hodnotu exponenciální funkce tak, jako ve vztahu (7).

Členy $\ln|\boldsymbol{\Sigma}_r|$ ve vztahu (25) a $1\Big/\sqrt{(2\pi)^N|\boldsymbol{\Sigma}_r|}$ ve vztahu (7) jsou nezávislé na klasifikovaném vektoru $\boldsymbol{x}$, můžeme je vypočítat před vlastní klasifikací jako konstantu pro každou třídu zvlášť a na výpočetní náročnosti algoritmů se tedy neprojeví.

5 Implementace

Oba klasifikační algoritmy byly implementovány ve výpočetním systému MATLAB.

6 Výsledky

Jak již bylo uvedeno, oba klasifikační algoritmy jsou co se týče výsledků klasifikace totožné a nemá smysl je z tohoto pohledu jakkoli porovnávat.

Zajímavé však může být srovnání výpočetní náročnosti klasifikátorů postavených na těchto algoritmech, kde se ukazuje, že klasifikátor založený na diskriminačních funkcích je na stejném hardwaru zhruba o 20 % rychlejší, než Bayesův klasifikátor. Rychlost obou klasifikátorů byla ověřena na sadě 94550 příznakových vektorů.

7 Závěr

Bayesův klasifikátor i klasifikátor využívající diskriminačních funkcí jsou klasifikátory založené na statistické analýze dat. Oba klasifikátory poskytují dostatečně spolehlivé výsledky (přesnost klasifikace byla na použitých datových souborech větší než 95 %) a jsou snadno implementovatelné v prostředí MATLAB. Jak bylo ukázáno, použití diskriminačních funkcí může být výhodnější s ohledem na menší výpočetní náročnost klasifikátoru.

Poděkování

Tato práce byla podporována výzkumným záměrem č. MSM 6840770012 (udělen Ministerstvem školství, mládeže a tělovýchovy České republiky).

Reference

[1] Jan Havlík and Zdeněk Horčík. Bayes classification of thumb motion. In *5th European Symposium on Biomedical Engineering ESBME'06 [CD-ROM]*, 2006.

[2] Jan Havlík, Jan Uhlíř, and Zdeněk Horčík. Thumb motion classification using discrimination functions (in press). In *Applied Electronics 2006*. University of West Bohemia in Pilsen, Pilsen, 2006.

[3] Zdeněk Kotek, Vladimír Mařík, Václav Hlaváč, Josef Psutka, and Zdeněk Zdráhal. *Metody rozpoznávání a jejich aplikace*. Academia, Praha, 1993.

[4] Karel Rektorys et al. *Přehled užité matematiky II*. Prometheus, Praha, 2000.

Jan Havlík
Katedra teorie obvodů, Fakulta elektrotechnická, České vysoké učení technické v Praze
Technická 2, 166 27 Praha 6, tel: 224 352 048, email: xhavlikj@fel.cvut.cz