

UŽITÍ KOHERENČNÍ FUNKCE PRO DISTRIBUOVANOU ANALÝZU VÍCEKANÁLOVÝCH SIGNÁLŮ

Robert Háva, Aleš Procházka

Vysoká škola chemicko-technologická, Ústav počítačové a řídicí techniky

Abstrakt

Analýza vícekanálových signálů je významná v celé řadě aplikačních oblastí. Příspěvek se přitom omezuje na užití obecných matematických metod pro zpracování paralelních časových řad EEG záznamů s užitím koherenční funkce a implementací distribuovaného počítání pro optimalizaci příslušného výpočetního času. Těžiště práce je v odvození algoritmu umožňujícího paralelní zpracování dat na více výpočetních systémech a posouzení dosažených výsledků.

1 Úvod

Analýza elektroencefalografických (EEG) záznamů je předmětem zájmu řady prací [3, 4], které se zaměřují na vztah mezi dílčími kanály, na segmentaci a klasifikaci záznamů i sledování vztahu mezi časovým vývojem EEG záznamů a obrazy magnetické resonance. Příspěvek se omezuje na analýzu pozorovaných časových řad a jejich zpracování s využitím vlastních a knihovních funkcí.

Významná část práce je věnována možnostem paralelního výpočtu na vícejádrových procesorech a počítačích propojených v počítačové síti s využitím příslušných knihoven MATLABu. Výsledkem studie je posouzení možností paralelního výpočtu.

2 Koherenční funkce

Koherenční funkce [3] je často používanou charakteristikou pro popis vztahu mezi frekvenčními vlastnostmi časových řad. Hodnoty této funkce leží v intervalu $\langle 0, 1 \rangle$ a vypovídají o podobnosti signálů v jednotlivých frekvenčních pásmech. Vlastní definice této funkce vychází z definice vlastní a vzájemné výkonové spektrální hustoty podle vztahu

$$C_{xy}(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)}, \quad 0 \leq C_{xy}(f) \leq 1 \quad (1)$$

pro frekvence $f \in \langle 0, f_s/2 \rangle$, kde značí

$S_{xy}(f)$ - odhad vzájemné výkonové spektrální hustoty

$S_{xx}(f)$ - odhad vlastní výkonové spektrální hustoty pro signál $\{x(n)\}$

$S_{yy}(f)$ - odhad vlastní výkonové spektrální hustoty pro signál $\{y(n)\}$

f_s - vzorkovací frekvenci

Koherenční funkce může nabývat hodnot v rozmezí $\langle 0, 1 \rangle$, přičemž hodnoty blízké 0 vypovídají o nekorelovanosti příslušných frekvenčních složek a hodnoty blízké 1 svědčí o jejich souvislosti. Správný odhad výkonové spektrální hustoty stacionárního signálu $\{x(n)\}_{n=0}^{N-1}$ a $\{y(n)\}_{n=0}^{N-1}$ lze provést na základě průměrování periodogramů určených z daného počtu L realizací procesu. Pro případ vzájemné výkonové spektrální hustoty platí

$$S_{xy}(f) = \frac{1}{L} \sum_{k=1}^L X_k^*(f) Y_k(f) \quad (2)$$

kde $X_k(f)$ a $Y_k(f)$ jsou spektra dílčích signálů určená pomocí jejich diskrétní Fourierovy transformace (DFT). Dosazení vztahu (2) do definičního vztahu (1) dostaneme odhad koherenční funkce [2] ve tvaru

$$C_{xy}(f) = \frac{\left| \sum_{k=1}^L X_k^*(f) Y_k(f) \right|^2}{\sum_{k=1}^L X_k^*(f) X_k(f) \sum_{k=1}^L Y_k^*(f) Y_k(f)} \quad (3)$$

a tedy

$$C_{xy}(f) = \frac{\left| \sum_{k=1}^L X_k^*(f) Y_k(f) \right|^2}{\sum_{k=1}^L |X_k(f)|^2 \sum_{k=1}^L |Y_k(f)|^2} \quad (4)$$

Z hlediska algoritmizace výpočtu je podstatný vztah (4), ze kterého plyne, že je nutné

- průměrování komplexního spektra v čitateli a následný výpočet čtverce modulu
- průměrování čtverců modulů ve jmenovateli

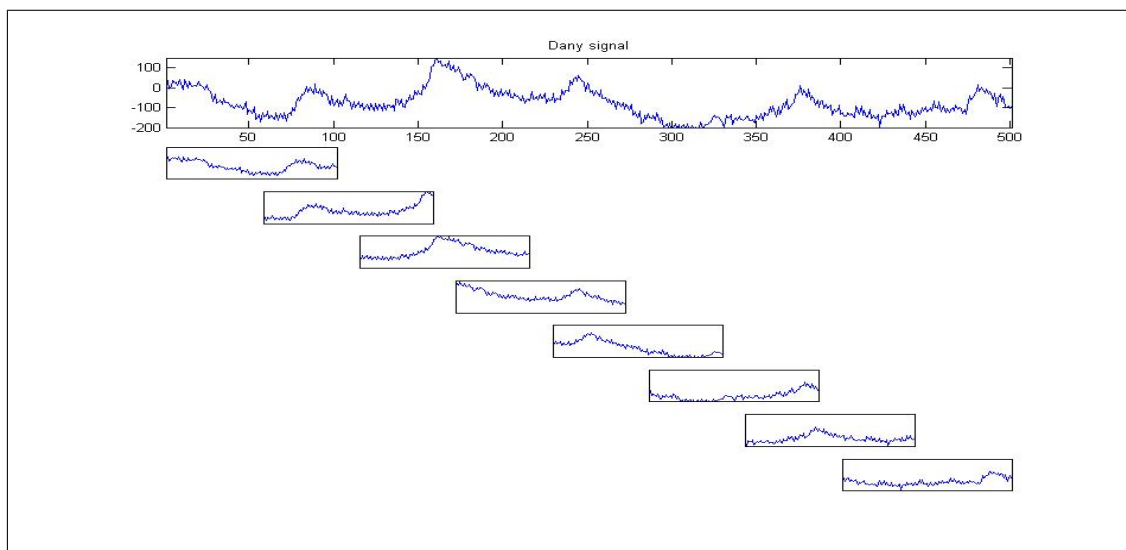
V případě realizace o konečné délce, lze provést její rozdělení na L úseků stejné délky s případným překrýváním dílčích úseků a následně aplikovat průměrování odhadů výkonových spektrálních hustot.

3 Tvorba koherenční funkce

V prostředí výpočetního systému MATLAB lze pro výpočet hodnot koherenční funkce užít standardní funkci *mscohere*, která pro dané posloupnosti $\{x(n)\}_{n=0}^{N-1}$ a $\{y(n)\}_{n=0}^{N-1}$ stanoví její odhad pomocí vztahu

```
Cxy=mscohere(x,y)
```

Pro posouzení vlastností koherenční funkce byl sestaven vlastní výpočetní algoritmus se stejnými vstupními parametry, který poskytuje výsledky shodné s funkcí *mscohere*. Toto vlastní programové řešení vychází z matematické definice popsané v předešlé části, poskytuje však možnosti podrobnější analýzy této funkce a její implementace v rámci distribuovaného počítání. Celá realizace $\{x(n)\}_{n=0}^{N-1}$ procesu je pro potřeby výpočtu rozdělena na $L=8$ oken, které se z 50 % překrývají podle obr. 1.



Obrázek 1: Dělení signálu na L částí pro potřeby výpočtu koherenční funkce

Počet překrývajících se hodnot je roven počtu $S=N/(L+1)$ a dílčí okna obsahují $M = 2 S$ hodnot. Navržený algoritmus zahrnující příkazy pro zpracování dat ve vlastní grafickém uživatelském prostředí (GUI) je uveden na obr. 2.

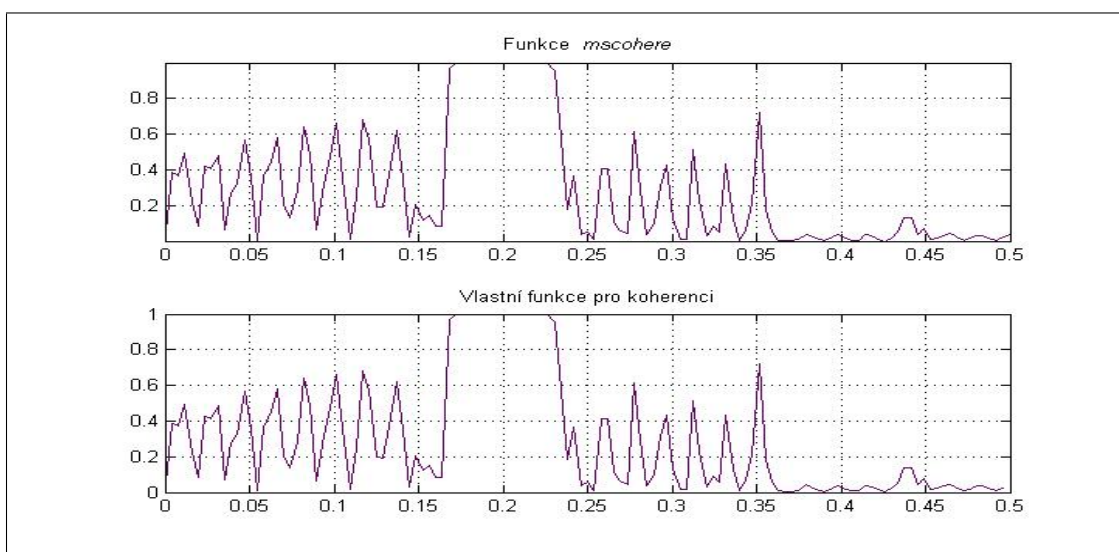
```
L=8; F=256;
x=handles.sig1;y=handles.sig2;
N=length(x);
S=floor(N/(L+1)); M=S*2;
for i = 1:L
    X(:,i)=x((i-1)*S+1:(i-1)*S+M).*hamming(M);
    Y(:,i)=y((i-1)*S+1:(i-1)*S+M).*hamming(M);
end
XX=fft(X,F); YY=fft(Y,F);
Sxy=XX.*conj(YY); % cross spectral density
Sxx=abs(XX).^2; % auto spectral density of x
Syy=abs(YY).^2; % auto spectral density of y
Cxy=abs(mean(Sxy,2)).^2./(mean(Sxx,2).*mean(Syy,2));
```

Obrázek 2: Algoritmus pro výpočet koherenční funkce, kde značí: L - počet oken pro dělení signálu, F - počet hodnot diskretní Fourierovy transformace, x, y - signály, *hamming* - název funkce pro Hammingovo okénko, *fft* - funkci pro diskretní Fourierovu transformaci, *mean* - funkci pro výpočet střední hodnoty, Cxy - hodnoty koherenční funkce

Pro posouzení správnosti algoritmu byly analyzovány stejné posloupnosti pomocí funkce *mscohere* i pomocí vlastní funkce s výsledkem na obr. 4. Délka potřebného výpočetního času při standardním a distribuovaném počítání v systému MATLAB byla přitom detekována pomocí funkcí uvedených na obr. 3 s výsledky v tabulce 1.

```
tic % zacatek mereni casu
Cxy=mscohere(x,y)
toc % konec mereni casu
```

Obrázek 3: Výpočet doby výpočtu



Obrázek 4: Porovnání koherenčních funkcí určených standardním a vlastním programem

Tabulka 1: Porovnání výpočetního času

<i>Funkce</i>	Čas (jeden výpočet)	Čas (300 iterací)
mcohore	0.2413	61.7623
vlastní funkce	0.1191	34.8863

4 Filtrace dat

Pozorovaný EEG signál je ve většině případů významně znehodnocen aditivní rušivou složkou o síťové frekvenci 50Hz a pro další analýzu je nutné její potlačení této. V případě užití FIR filtru lze jeho konstanty $\{b_k\}_{k=0}^{M-1}$ určit standardními příkazy MATLABu ve tvaru

```
M=101; fs=128; fc1=45; fc2=55; cut=[fc1 fc2]/(fs/2);
b1=fir1(M-1, cut, 'stop');
```

kde značí

M - řád filtru

$fc1, fc2$ - mezní frekvence pásmové zadržky (specikované parametrem 'stop' ve funkci fir1)

fs - vzorkovací frekvenci

Po vypočtení konstant $\{b_k\}_{k=0}^{M-1}$ lze určit filtrovanou posloupnost hodnot k dané posloupnosti $\{x_n\}_{n=0}^{N-1}$ vztahem

$$EEGden(n) = b_0 x(n) + b_1 x(n-1) + \dots + b_{M-1} x(n-M+1) \quad (5)$$

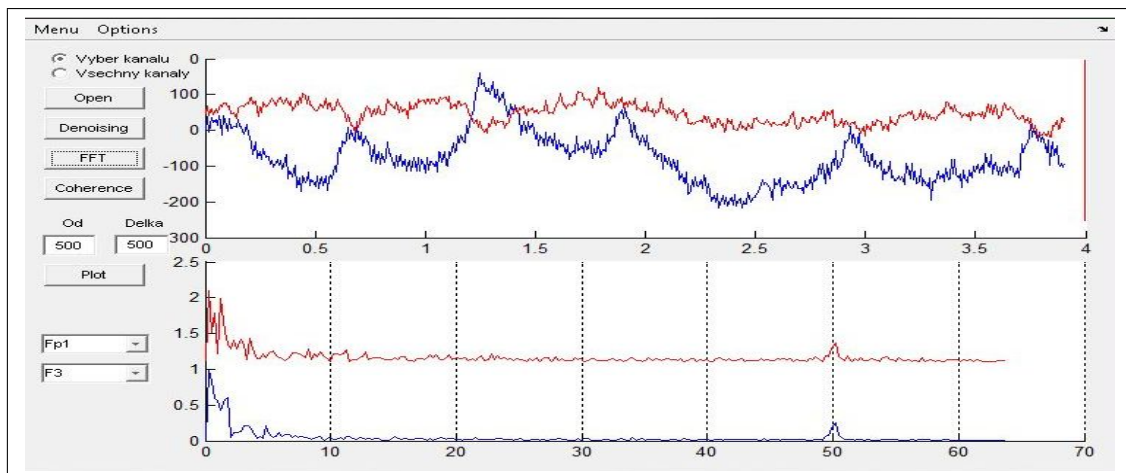
s využitím funkce

```
EEGden=filter(b1,1,signal);
```

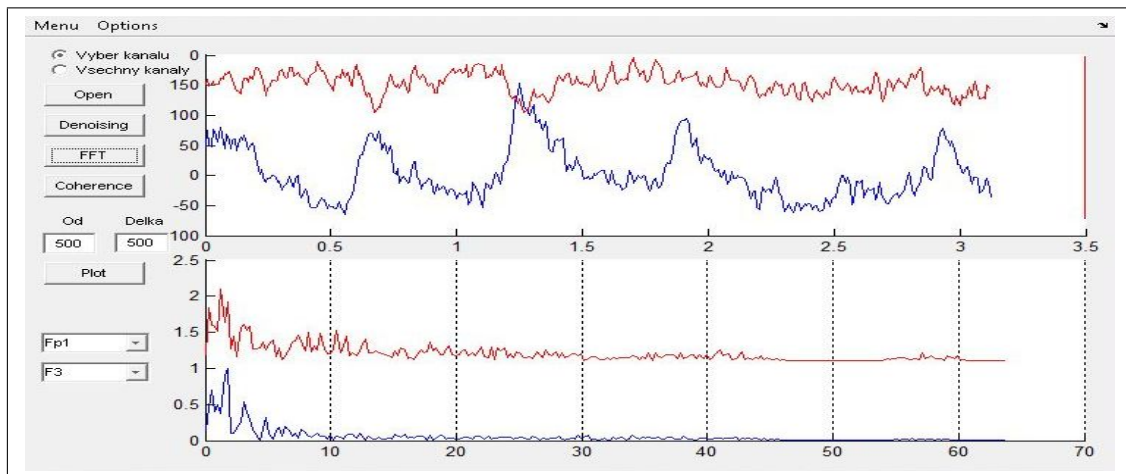
V případě analýzy EEG signálu se nejprve realizuje potlačení síťové frekvence 50Hz pásmovou zadržkou a následný výběr frekvenčních složek v pásmu $\langle 0.5, 60 \rangle \text{Hz}$ pomocí programového úseku uvedeného na obr. 5.

```
M=101; fs=128; fc1=45; fc2=55; cut1=[fc1 fc2]/(fs/2);
M=101; fc1=0.5; fc2=60; cut2=[fc1 fc2]/(fs/2);
b1=fir1(M-1, cut1, 'stop'); b1=b1'; b2=fir1(M-1, cut2);
b2=b2';
EEGdenx=filter(b1,1,signal);signal=filter(b2,1,EEGdenx);
```

Obrázek 5: Úsek programu pro číslicovou filtraci daných časových řad



Obrázek 6: Grafické uživatelské prostředí pro znázornění průběhu zvolených úseků dvou vybraných kanálů EEG záznamu a jejich spekter



Obrázek 7: Grafické uživatelské prostředí pro znázornění průběhu zvolených úseků dvou vybraných kanálů EEG záznamu a jejich spekter po číslicové filtraci

5 Funkce distribuovaného počítání

Distribuované počítání [1] urychluje výpočet jednoduchým způsobem tak, že se složitější výpočty rozdělí na více menších úkolů a ty se odesílají počítačům připojeným k serveru. Počítače distribuované na počítačové síti následně zpracovávají dané úkoly a vrací zpět výsledky podle schématu na obr. 5, ve kterém značí

Job - úkol, který se posílá na plánovací server rozdělující úkoly

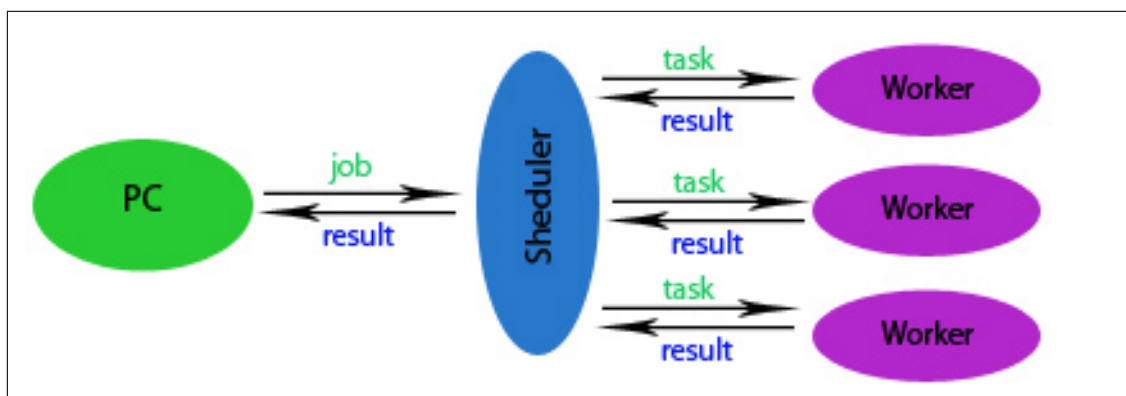
Task - konkrétní úkol, který se počítá na daném počítači

Result - výsledek, který se vrací zpět na počítač

Scheduler - plánovací server, který rozděljuje výpočty

Worker - počítač, který počítá jednotlivé úkoly

Na obrázku je vidět princip distribuovaného počítání, ale i jeho problém spočívající v nutnosti přenosu velkého objemu informací. Pokud je nutno pomocí distribuovaného počítání realizovat větší počet jednoduchých výpočtů, může dojít k tomu, že výpočet prováděný tímto způsobem může trvat déle než na jednom počítači. Důvodem je přitom nutnost komunikace počítačů mezi sebou po počítačové síti a posílání dat po každém výpočtu. Pokud je nutné počítat více jednoduchých úloh pomocí distribuované počítání, lze doporučit algoritmus, ve kterém se jako jeden TASK posílá několik výpočtů, při kterých se využijí počítače po delší dobu.



Obrázek 8: Princip distribuovaného počítání

6 Programové řešení v MATLABu

Matlab má pro distribuované počítání vlastní knihovnu (*Distributed Computing Toolbox*), která zahrnuje jak *Jobmanager*, tak *Scheduler*. Definice pro *Jobmanagera* je zajištěna příkazem

```
jm = findResource('jobmanager', 'name', 'MyJobManager');
```

Následně je nutné vytvořit konkrétní *Job* příkazem

```
j = createJob(jm); % konkrétní job
```

a definovat konkrétní *Task* pro vlastní výpočet příkazem

```
createTask(job, @EEGtask1, 4, {signal}); % tvorba tasku(ukolu)
```

Příkaz

```
submit(j);
```

zařadí *Task* do fronty a v okamžiku, kdy je volný *Worker*, provede jeho zpracování. Zpětný přenos výsledků z pracovních počítačů je následně zajištěn příkazem

```
waitForState(j, 'finished'); % čekání než se dopocita job
```

```
results = getAllOutputArguments(j) % získání výsledku z workru
```

7 Rychlost výpočtu

Rychlost výpočtu lze ovlivnit buď lepší optimalizací výpočetní metody a nebo počítáním na výkonnějších počítačích. V tabulce 2 je porovnána rychlost výpočtu na systémech s jedním nebo více výpočetními jádry procesoru v případě různých operačních systémů.

Tabulka 2: Tabulka rychlostí výpočtu

Platforma	Počet jader	Čas výpočtu
Vista 32bit	1	37.81
Vista 32bit	4	31.47
Vista 64bit	1	27.00
Vista 64bit	2	22.68
distribuované počítání	–	17.80

8 Závěr

Další práce bude věnována posouzení vztahu mezi dílčími kanály EEG záznamu pro rozsáhlý soubor vstupních dat. Následné studie se budou věnovat analýze hlavních komponent a segmentaci pozorovaných signálů.

Práce byla vytvořena v rámci výzkumného záměru MSM 6046137306 Fakulty chemicko-inženýrské Vysoké školy chemicko-technologické v Praze.

Literatura

- [1] *Distributed Computing*. <http://www.mathworks.com>.
- [2] Pavel Sovka Jan Uhlíř. *Číslíkové zpracování signálů*. Vydavatelství ČVUT, 2002.
- [3] Saeid Sanei and J. A. Chambers. *EEG Signal Processing*. Wiley - Interscience, 2007.
- [4] Saeed Vaseghi. *Advanced Digital Signal Processing and Noise Reduction*. John Wiley and Sons Ltd., 2006.

Bc. Robert Háva, Prof. Aleš Procházka
Vysoká škola chemicko-technologická v Praze
Ústav počítačové a řídicí techniky
Technická 1905, 166 28 Praha 6
Telefon: 00420-220 444 198 * Fax: 00420-220 445 053
E-mail: Robert.Hava@vscht.cz, A.Prochazka@ieee.org