

HILBERT-HUANGOVA TRANSFORMACE APLIKOVANÁ V EXPERTNÍM SYSTÉMU

Application of Hilbert-Huang Transform in an Expert System

Josef Kokeš

Abstrakt: Paper describes one of successful attempts to utilize Hilbert-Huang Transform in an expert system. Expert system for automatic unattended trading on american share exchanges was developed. Basic principles and selected results are shown in this article.

Key words: Hilbert-Huang Transform, Expert Systém, Intrinsic function, sifting, broker, share.

1. Cíl a prostředky

Základním cílem mého výzkumu proto bylo, vytvořit expertní systém použitelný v reálném čase. Byl zvolen princip *multithreadové* aplikace, jak jsem o tom referoval v [3]. Jakmile byla vyřešena otázka bezpečného paralelního zpracování mnoha pravidel v expertním systému, byla už další realizace poměrně jednoduchá a přehledná. S výhodou bylo využito skutečnosti, že pravidla v expertním systému mohou obsahovat libovolnou procedurální znalost, a tak bylo nejprve vytvořeno několik „obslužných“ pravidel.

V článku popisují jeden pokus o nasazení expertního systému v praxi. Je však třeba poctivě poznamenat, že vedle tohoto úspěšného pokusu jsem absolvoval ještě řadu dalších pokusů neúspěšných. Celková výtěžnost proto byla blízká nule, nebo spíše záporná.

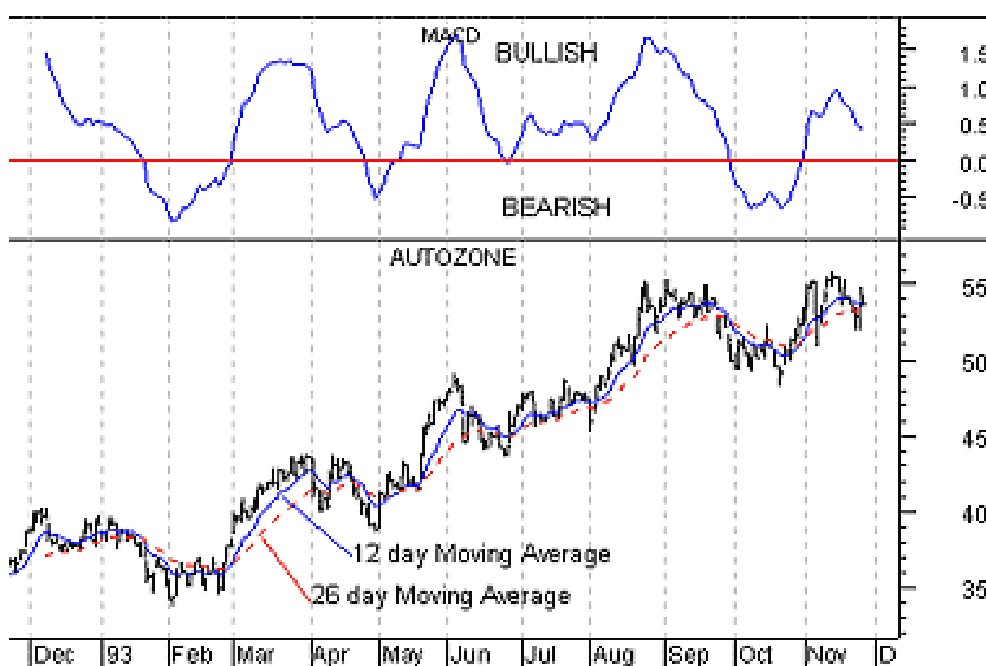
V tomto konkrétním případě jsem využil služeb internetové brokerské společnosti nyní TD AmeriTrade a na její server jsem se připojil. Pravidlo *VstupDat* bylo napsáno tak, že simulovalo chování běžného MS Internet Exploreru. Proud dat, posílaný serverem společnosti DATEK, byl pravidlem *VstupDat* zachycen, očištěn od nepodstatných informací a ve vhodné formě uložen pro potřeby ostatních pravidel. Pravidlo má čistě servisní charakter, k žádnému logickému zpracování informací v tomto pravidle nedochází až na to, že agreguje všechny obchody realizované v dané sekundě.

Zcela obdobně, pravidlo *Rizeni* se stará o provedení příkazů k obchodům. To znamená, že vůči serveru DATEK simuluje chování běžného MS Internet Exploreru tak, jako kdyby příkazy k provedení obchodů prováděl člověk na klávesnici počítače.

Dalším zajímavým a důležitým pravidlem je *WatchDog*. Toto pravidlo v podstatě obsahuje časovač a jeho úkolem je hlídat, jestli se některá část programu nezastavila. Zejména

u internetového spojení se totiž často stávalo, že se přenos z nějakého důvodu zastavil. Pravidlo *WatchDog* fakticky funguje stejně, jako když si uživatel vyžádá překreslení www stránky klávesou F5. Pravidel typu *WatchDog* bylo v různých částech aplikace realizováno několik. Pravděpodobně by bylo vhodné, mít též jakýsi generální *WatchDog*, který by hlídal funkčnost celé aplikace na globální úrovni. Touto možností jsem se však nezabýval.

Všechna obslužná pravidla běží v paralelních *threadech*, tedy nezávisle na sobě navzájem i na zbytku inferenční sítě.



Obrázek 1 Typický průběh ceny akcie vykazuje určitou periodicitu. Nahoře průběh indikátoru MACD.

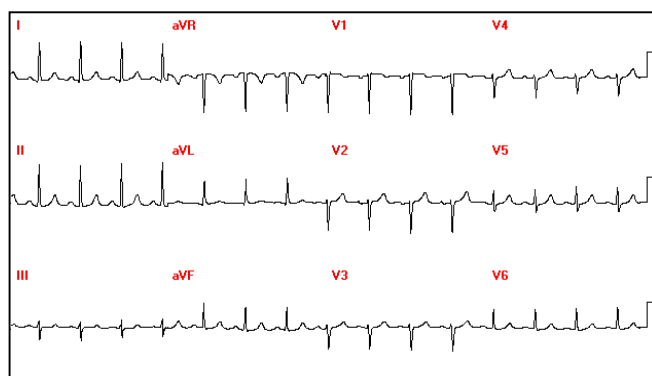
Vlastní odvozovací pravidlová síť expertního systému byla v prvních pokusech realizována jako velice jednoduchá. K rozhodování o okamžicích provedení obchodu bylo použito pravidlo, známé v burzovní terminologii jako indikátor MACD. Jednoduchý příklad tohoto pravidla ukazuje obrázek. Na něm je indikátor MACD, který se vypočítá odečtením dlouhodobějšího (zde 26 dní) klouzavého průměru ceny od krátkodobého (zde 12 dní) klouzavého průměru. Subjektivní míra důvěry tohoto pravidla byla nastavena jako vysoká, takže se fakticky jednalo o hlavní pravidlo, na základě kterého expertní systém rozhodoval o zahájení obchodu. Vedle něj se uplatňovala ještě řada méně významných rozhodovacích pravidel, jako například pravidla o celkovém trendu indexů burzy (indexy stoupají, indexy klesají) a o denním čase (ke konci obchodního dne je tendence pozice spíše uzavírat, na začátku obchodního dne spíše otevírat). Tato pravidla se však na rozhodování podílela spíše okrajově.

Brzy se ukázalo, že toto jednoduché pravidlo vyhovuje jen ve velmi specifické situaci burzy. Proto byl zahájen výzkum, zaměřený na studium nestacionárních procesů a popsany v následujících odstavcích.

2. Predikce nestacionárních dějů

Mnozí filosofové věří, že celý svět se vyvíjí v cyklech. Ať už je tomu jakkoli, v technické praxi se setkáváme s množstvím jevů, které výrazně cyklický charakter mají. Pro jejich popis používáme modely založené na periodicitě. Na obrázcích 1 až 3 jsou příklady naměřených dat; ve všech ukázaných případech cítíme, že jistý prvek cykličnosti v nich obsažen je. Kdybychom uměli nějakým způsobem nelineárně přetransformovat (podle potřeby někde zkrátit, jinde prodloužit), *morfovat*, časovou osu tak, aby všechny cykly získaly stejnou periodu, mohli bychom popsat jejich tvar a dokázali bychom **predikovat** jejich budoucí průběh.

Dá se ukázat, že zcela ekvivalentní je, časovou osu ponechat lineární a naopak podle potřeby průběžně měnit **okamžitou frekvenci** periodického průběhu. Tento postup se na první pohled zdá schůdnější, protože v technické praxi pomocí Fourierovy transformace běžně přecházíme mezi popisem v časové a ve frekvenční oblasti. Můžeme proto předpokládat, že pomocí Fourierovy transformace vytvoříme spektrum průběhu, v něm vyhledáme nejvýznamnější spektrální složku (např. složku nesoucí maximum energie) a tu prohlásíme za okamžitou frekvenci signálu v daném bodě. Je to jednoduché, srozumitelné – a nesprávné.

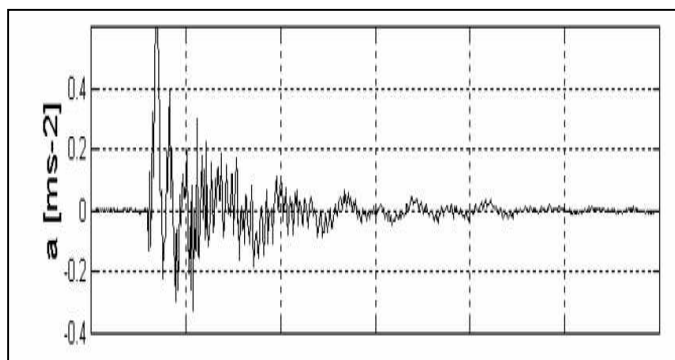


obrázek 2 Typický EKG záznam zdravého srdce



obrázek 3 Periodicita v ceně akcií

Chyba je v tom, že Fourierova transformace, vedle vcelku samozřejmých obecných předpokladů, vyžaduje, aby (a) systém byl lineární a (b) data byla stacionární. Stacionaritou zde míníme požadavek, aby kovarianční funkce byla invariantní vůči časovému posunu (jde o t.zv. stacionaritu 2. řádu). Žádný z těchto požadavků neumíme beze zbytku splnit. Zatímco s nelinearitou systému se umíme do jisté míry vyrovnat pomocí linearizace, požadavek stacionarity je v praxi nesplnitelný.



obrázek 4 Zrychlení naměřené čidlem na železničním svršku

To se v roce 1946 pokusil vyřešit Gabor tím, že ze signálu uvažoval vždy jen krátké časové okno, ve kterém předpokládal, že je průběh dostatečně stacionární. Podle potřeby pak toto okno posouval po naměřeném průběhu. To je princip krátkodobé Fourierovy transformace, STFT (*Short Time Fourier Transform*), někdy označované též jako *Windowed Fourier Transform*. Jestliže pro časově spojité signály je běžná Fourierova transformace definována vztahem

$$F\{x(t), \omega\} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} x(\tau) e^{-j\omega\tau} d\tau$$

pak krátkodobá Fourierova transformace STFT v čase t_0 je dána podobným vztahem

$$F\{x(t), \omega, t_0\} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} x(\tau) h(t_0 - \tau) e^{-j\omega\tau} d\tau$$

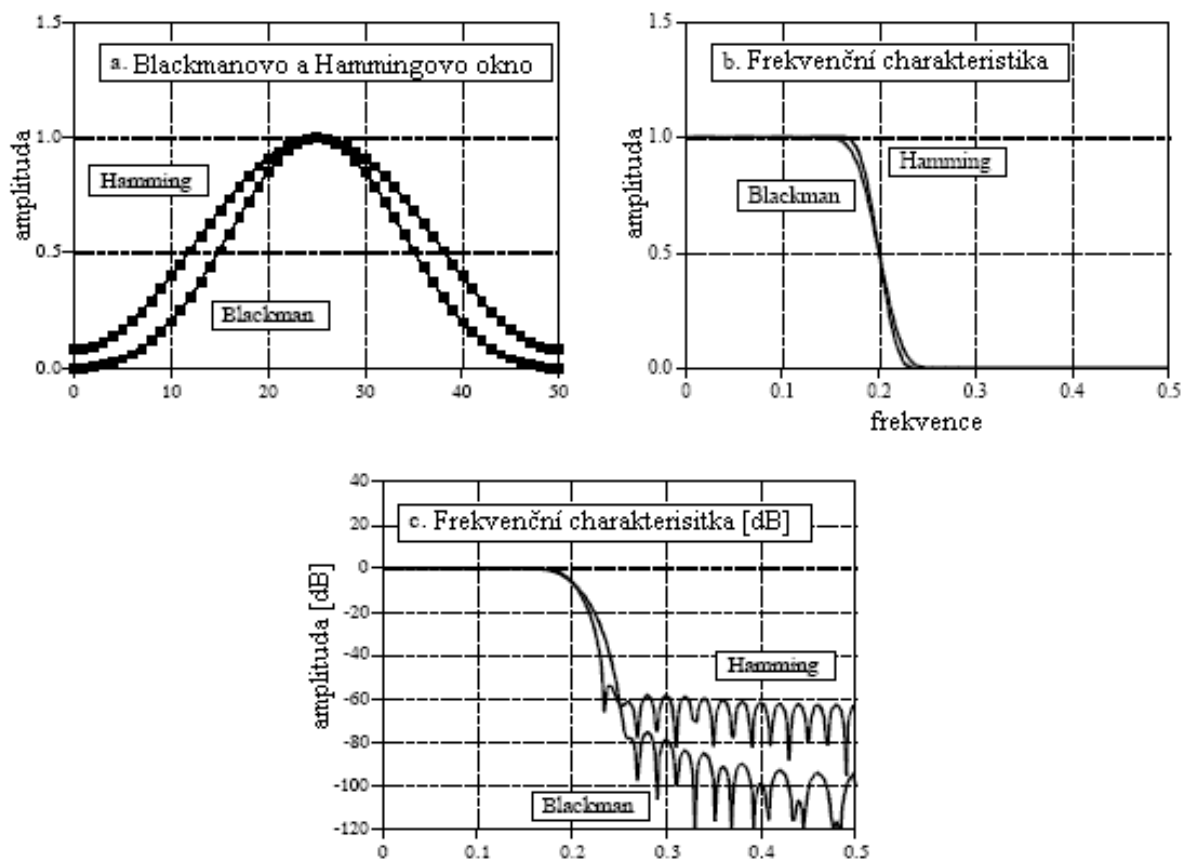
ve kterém $h(\tau)$ je reálná, symetrická funkce realizující časové okno, pro kterou z důvodu normalizace platí $\|h(\tau)\|=1$. V principu by $h(\tau)$ mohla být i obdélníková funkce, ale z praktických důvodů se dává přednost oknům, která ve spektru produkují menší rušivé efekty. K nejpoužívanějším patří „kosínusový zvon“ a z něj odvozené Hammingovo okno

$$h_{Hamming}(\tau) = 0,54 - 0,46 \cos(2\pi\tau)$$

nebo Blackmanovo okno

$$h_{\text{Blackman}}(\tau) = 0,4 - 0,5 \cos(2\pi\tau) + 0,08 \cos(4\pi\tau).$$

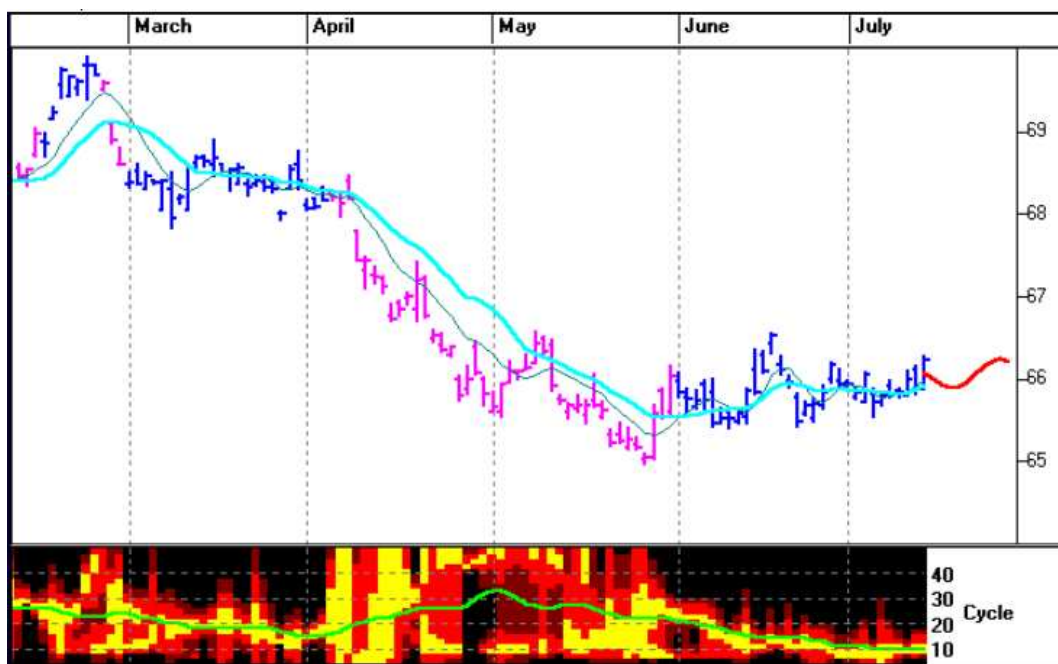
Na obrázku 4a je vidět, že klesání Hammingova okna je zhruba o 20% rychlejší než u Blackmanova. Avšak na části c) vidíme, že potlačení u Blackmanova okna je -74dB, zatímco u Hammingova okna je to pouze -53dB.



obrázek 5 Porovnání Blackmanova a Hammingova okna

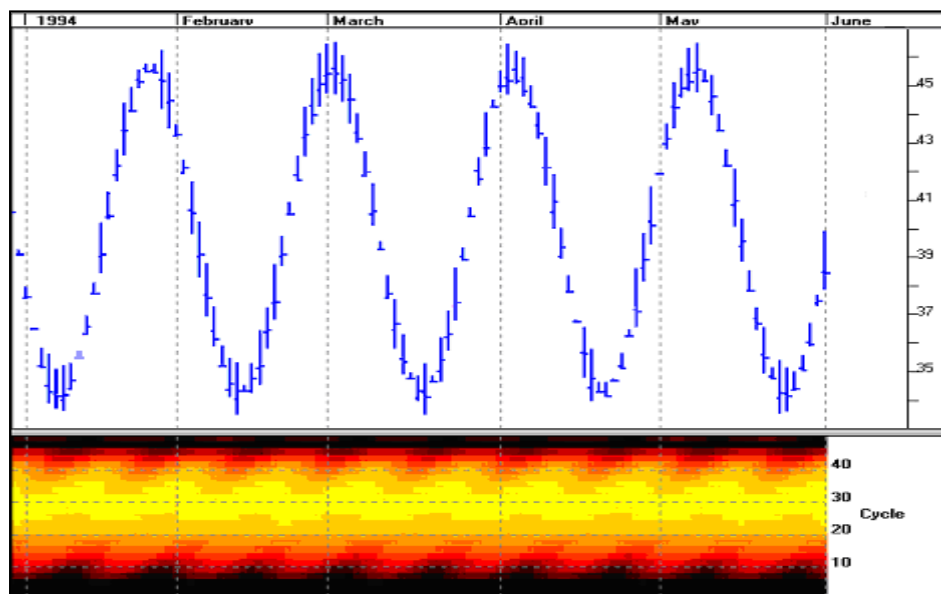
Okénko se postupně posouvá přes celý zkoumaný signál. Pro každý čas t_0 odpovídající středu okna tak dostaneme dílčí spektrum, které odpovídá okamžité poloze okna. Bývá zvykem, že se tato dílčí spektra pro časy t_0 vynášejí pod časovým průběhem, kde vytvářejí spektrogram. Přitom na vodorovné ose je čas (shodný s časem zkoumaného průběhu) a na

svislé ose je frekvence. Spektrální výkonová hustota („amplituda“ spektra) se vyjadřuje intenzitou zbarvení. Na obrázku 5 je příklad spektrogramu pro jednoduchý simulovaný průběh. Je jím sinusový signál o délce periody $T=24$ jednotek. Protože tento signál je stacionární, jsou všechna dílčí spektra stejná (v rámci přesnosti).



obrázek 6 Spektrogram sinusového signálu o periodě $T=24$ při šířce okna $W=64$

U signálů, které jsou sice nestacionární, ale mění se pomalu, je změna uvnitř okna malá a výsledky získané pomocí STFT jsou použitelné. Ale u rychle se měnících průběhů, jako jsou příklady na obrázcích 1 až 3, narazíme na problém: pokud zvolíme široké okno, průběh uvnitř okna nebude ani přibližně stacionární a dostaneme zkreslené výsledky. Naproti tomu, pokusíme-li se okno zúžit v časové oblasti, ve frekvenční oblasti úměrně klesne rozlišovací schopnost. Je to dobře vidět na obr. 5, kde je zvoleno relativně úzké okno (64 jednotek) vzhledem k periodě signálu (24 jednotek). Vidíme, že určit ze spektrogramu nejvýznamnější frekvenci signálu není snadné, a to ani u tohoto čistě harmonického průběhu, protože maximum spektrální hustoty je značně ploché (oblast vyplněná nejsvětlejší barvou je vysoká) a zasahuje frekvence odpovídající periodám od $T=20$ až do $T=40$. Pro průběhy, které nejsou stacionární, jsou výsledky ještě horší. Na připojeném obrázku 6 je ukázán spektrogram, tentokrát vytvořený ze skutečných (nestacionárních) dat. Přímé určení nejvýznamnější spektrální složky, resp. okamžité frekvence, je ze spektrogramu prakticky nemožné. Přesto existují metody, jak ji určit lze a ve spektrogramu na obr. 6 je vyznačena zelenou čarou.



obrázek 6 Spektrogram ze sinusovky

3. Okamžitá frekvence

Pojem okamžité frekvence jsem doposud používal intuitivně. Nyní si jej zavedeme přesněji. Vyjdeme z Hilbertovy transformace, což je konvoluční integrál funkce $x(t)$ a výrazu $\frac{1}{\pi}$, daný vztahem

$$H\{x(t), t\} = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau$$

Pro lepší porozumění, co vlastně transformace dělá, lze tento vztah přepsat pomocí operátoru konvoluce na

$$H\{x(t), t\} = \frac{1}{\pi} x(t) * \frac{1}{t}$$

a po Fourierově transformaci obou stran dostaneme

$$F\{H\{x(t), t\}\} = \frac{1}{\pi} \cdot F\{x(t)\} \cdot F\left\{\frac{1}{t}\right\} = \frac{1}{\pi} \cdot X(jf) \cdot j\pi(-\text{signum}(f)),$$

což lze interpretovat takto: Hilbertova transformace **ve frekvenční oblasti** pracuje tak, že složkám se zápornou frekvencí posouvá fázi o $+\frac{\pi}{2}$ a složkám s kladnou frekvencí o $-\frac{\pi}{2}$, přičemž v obou případech zachovává jejich amplitudu.

Vezměme nyní naměřená data, což je reálná funkce času $x(t)$. Nejprve vypočítáme Hilbertovu transformaci $x(t)$

$$y(t) = H\{x(t), t\} = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau.$$

Nyní vytvoříme zvláštní komplexní signál $z(t)$ tak, že jako jeho reálnou složku vezmeme $x(t)$ a jako jeho imaginární složku vezmeme $y(t)$ neboli Hilbertovu transformaci původního $x(t)$. Tento signál nazveme **analytickým signálem** a vyjádříme ho pomocí amplitudy a fáze:

$$z(t) = x(t) + jy(t) = A(t) + e^{j\theta(t)}$$

Amplitudu i fázi můžeme vyjádřit pomocí $x(t)$ a $y(t)$

$$A(t) = \sqrt{x^2(t) + y^2(t)}$$

$$\theta(t) = \arctan\left(\frac{y(t)}{x(t)}\right)$$

Nyní již konečně můžeme vypočítat **okamžitou frekvenci** (anglicky *instantaneous frequency*) jako časovou derivaci fáze:

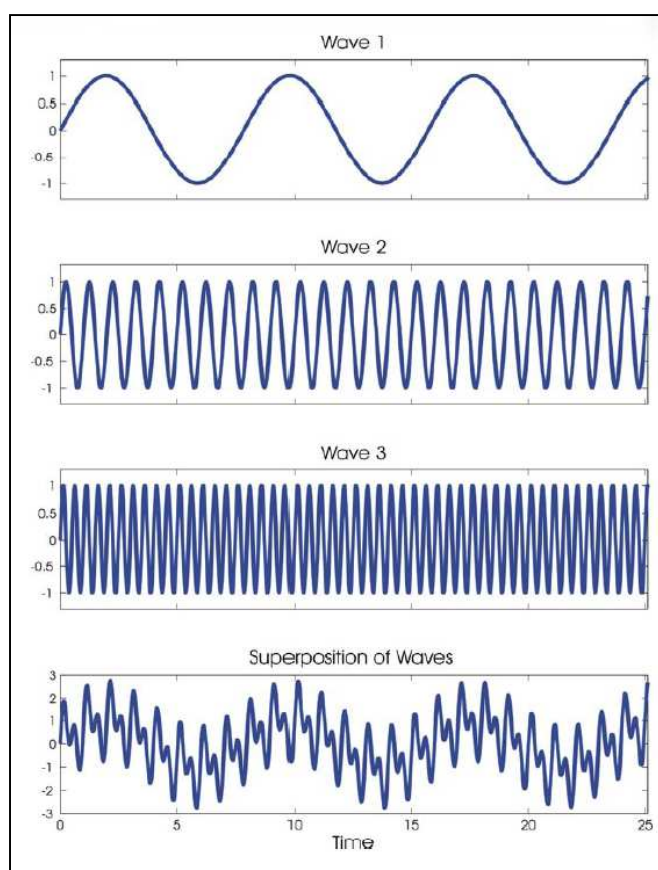
$$\omega_0 = 2\pi f_0 = \frac{d\theta(t)}{dt}.$$

4. Hilbert-Huangova transformace

Huangův přínos spočívá v tom, že Hilbertovu transformaci doplnil o algoritmus, proces, který nazval *sifting* (doslova „prosévání“). Pomocí prosévání je možno velkou třídu signálů dekomponovat na jejich jednodušší úzkopásmové složky. Tyto složky se nazývají **IMF** (*Intrinsic Mode Functions*). Český ekvivalent ještě není ustálen – snad by se hodilo „vlastní modální funkce“. Vlastní modální funkce je taková funkce, která má stejný počet lokálních extrémů jako původní průběh (také má počet průchodů nulou lišící se od něj maximálně o jedničku, aby vyhovovala tradiční podmínce pro Gaussovský úzkopásmový stacionární signál) a navíc střední hodnota mezi obálkami, počítaná přes všechny body, je

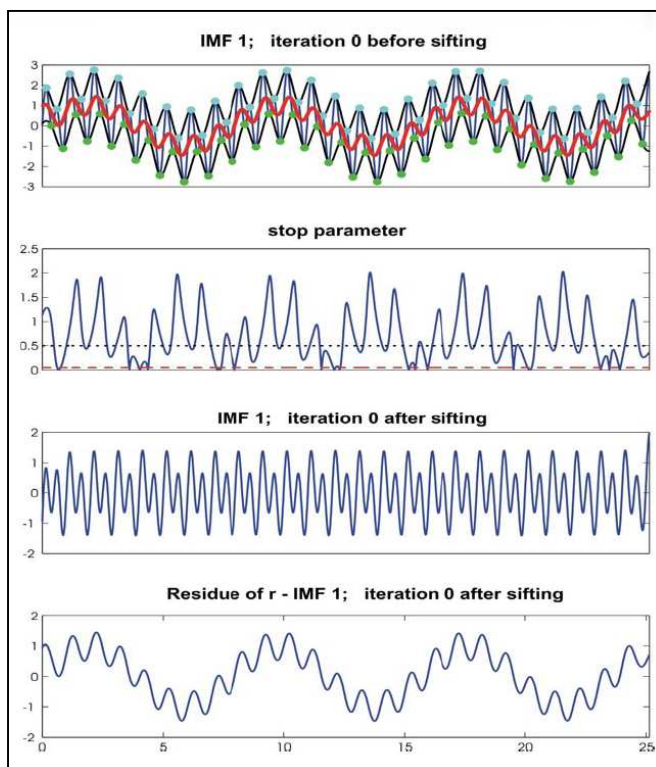
rovna nule. Pomocí **vlastních modálních funkcí** lze ze signálu vyextrahovat podstatné vlastnosti a použít je jak k popisu signálu, tak k jeho predikci.

Celý algoritmus prosévání si vysvětlíme na jednoduchém modelu. Pro jednoduchost vyjdeme z uměle vytvořeného signálu, který sestavíme superpozicí tří sinusovek o známých (nikoliv nutně soudělných) frekvencích. Příklad takového signálu je na obr. 7. Tento signál nyní budeme prosívat pomocí Huang-Hilbertovy transformace (**HHT**). Prvním krokem je zjištění řídicí vlny a symetrizace vlnového profilu.



obrázek 7 Výchozí signál vznikl superpozicí tří sinusovek

Nejprve nalezneme všechny lokální extrémů funkce $x(t)$. Na obrázku 8a jsou lokální maxima vyznačena světle modrými a lokální minima zelenými body. Přes všechna lokální maxima vytvoříme obálku maxim. Vytvoříme ji tak, že jednotlivými lokálními extrémů $e_{\max}(t)$ proložíme kubický splajn. Podobně vytvoříme obálku minim tak, že kubický splajn proložíme lokálními minimy. Obě obálky jsou na obrázku 8a vyznačeny tmavě modrou křivkou.



obrázek 8 První krok HHT

Nyní provedeme „narovnání“. Cílem je, dosáhnout takového stavu, aby v každém bodě byla střední hodnota mezi oběma obálkami co nejbližší nule. Dosáhneme toho tak, že vypočteme střední hodnotu mezi oběma extrémy $m(t) = \frac{e_{\max}(t) + e_{\min}(t)}{2}$ a o tuto střední hodnotu zmenšíme $x(t)$:

$$d(t) = x(t) - m(t)$$

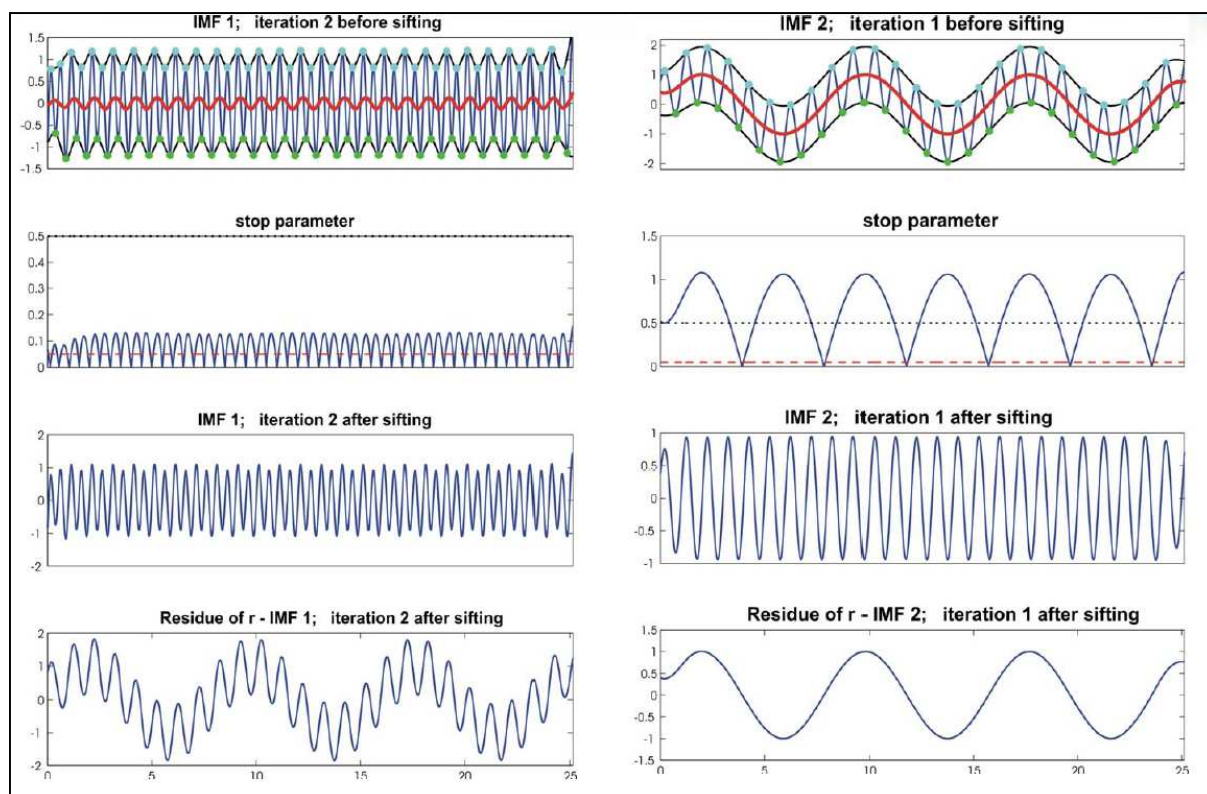
Tím jsem dostali první přiblížení k vlastní modální funkci prvního řádu, kterou označíme IMF1.

Prosívání je iterativní proces, proto celý postup (počínaje nalezením lokálních extrémů) budeme opakovat, dokud není splněno prosívací kritérium. Tímto kritériem je standardní odchylka mezi dvěma po sobě jdoucími iteracemi:

$$SD = \sum_{t=0}^T \left[\frac{|(h_{1(k-1)}(t) - h_{1(k)}(t))|^2}{h_{1(k-1)}^2} \right]$$

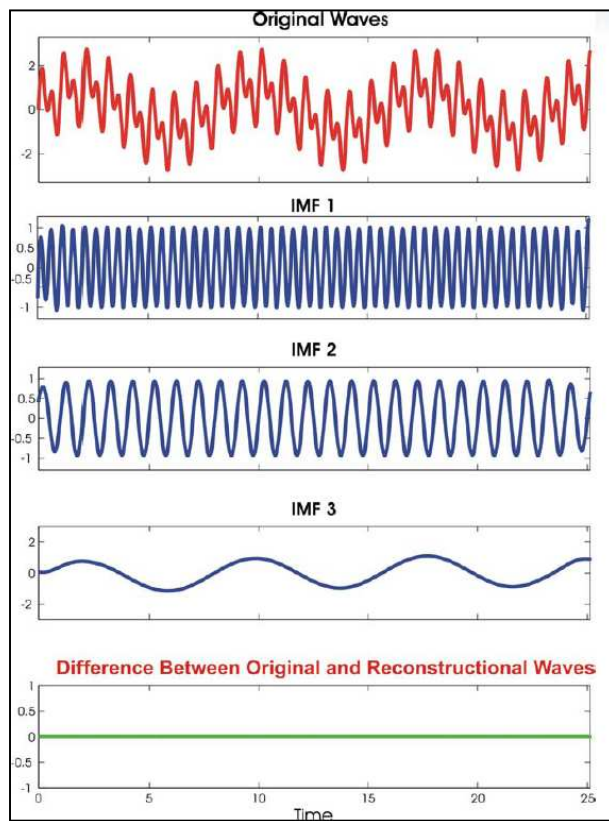
Na obrázku 8b je prosívací kritérium a na obr. 8c je stav po nulté iteraci a na obr. 8d je zbytek (reziduum) mezi IMF1 a původním signálem. Vidíme, že (obr. 8c) prosívací kritérium ještě není minimální, a proto následuje další iterace. Ta je zobrazena na levé polovině obrázku 9. Výsledkem celého iterativního procesu je upřesněná funkce IMF1 tak, jak ji vidíme na obr. 9 v levé polovině na pozici c) a dále zbytkové reziduum.

Toto reziduum je výchozím signálem pro výpočet IMF2. Druhá vlastní modální funkce se vypočítá stejným postupem, který začíná nalezením lokálních extrémů, postupuje iteracemi a končí splněním iteračního kritéria. Tento postup je ilustrován v pravé části obrázku 9.



obrázek 9 Iterativní výpočet vlastních modálních funkcí

Je zřejmé, že analogicky lze celý proces opakovat pro další IMF. Na obrázku 10 vidíme, že algoritmem HHT jsem dostali tři IMF, které se velmi málo liší od původních sinusovek, ze kterých jsem signál sestavili. Podstatné je, že jsou dobře periodické a stacionární, proto je můžeme – pokud to je potřeba - bez potíží identifikovat a použít k predikci.

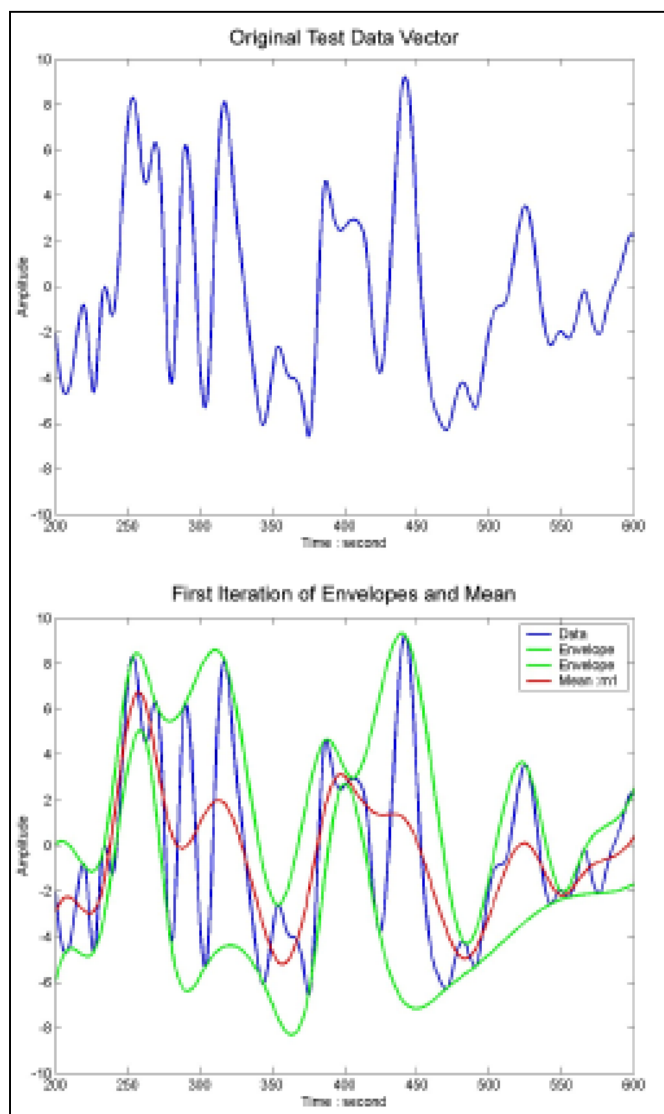


obrázek 10 Přehled výsledků HHT

Praktická aplikace tohoto algoritmu naráží na drobné těžkosti zejména na okrajích datového intervalu, kde kubické splajny jde těžko nastavit, protože mají tendenci divergovat mimo očekávanou oblast signálu.

5. Poznámky k významu HHT

Není bez jisté zajímavosti, že Huang, autor transformace HHT, je Číňan naturalizovaný v USA a pracující pro NASA. Samotná transformace HHT a s ní spojený algoritmus jsou pokládány za velmi perspektivní, a to dokonce do té míry, že NASA si algoritmus HHT nechala patentovat. Proto v otevřených zdrojích ani na internetu není snadno dostupný a za jeho využití (samozřejmě jen v dosahu US patentového zákona) vybírá NASA poplatky. Přesto jeho využití velmi rychle roste: jsou známé aplikace ze všech vědních oborů, a to od zpracování sonarových signálů námořnictvem (samozřejmě, že k detekci rybích hejn), přes predikci sluneční aktivity a zpracování údajů geomagnetických bouří, četné aplikace v medicíně, až po nasazení na akciových trzích. V roce 2002 byla HHT oceněna jako algoritmus roku 2002.



obrázek 11 Aplikace HHT na reálná data

6. Výsledky dosažené s první verzí ES

Expertní systém v první verzi byl testován v reálném provozu na americké akciové burze. K dispozici je autentický záznam obchodů z prvních pokusů s popisovaným expertním systémem tak, jak je realizoval on-line internetový broker DATEK .

V první řadě je zřejmé, že expertní systém neprováděl obchody každý den. Zčásti je to proto, že každý den není na burzách vhodná situace pro tento typ obchodů, zčásti (a hlavně) proto, že v mezidobích se na expertním systému pracovalo. Celkem v období od 16.10. do 12.11., tedy za 27 kalendářních dní, expertní systém obchodoval 11 dní.

Celkově expertní systém za 11 obchodních dní vytvořil čistý zisk 3 939,39 USD. Přitom objem, který byl investován, činil 20 000 USD. (Přesněji řečeno, investovány byly jen dvě třetiny z této částky, protože zbývající třetina představuje „margin“, což je účelová půjčka brokera. Protože otázka *marginu* je složitá a zde není prostor na její podrobnější vysvětlení, zanedbávám tuto okolnost a výpočty provádím, jako kdyby velikost investice byla 20 000. Mějme však na paměti, že fakticky jsou výsledky až o polovinu lepší, než zde uvedeno.)

Počet kalendářních dní	27
- z toho dní, kdy se obchodovalo	11
Nejlepší denní výsledek	2 396 USD
Nejhorší denní výsledek	-261 USD
Celkem zisk	3 939,39 USD
Zisk přepočtený na % z investovaných peněz:	
- za dny, kdy se obchodovalo	1,79% denně
- což v extrapolaci na 365,25 dní činí	654% p.a.
- za kalendářní dny	0,7295% denně
- což v extrapolaci na kalendářní rok činí	266,5% p.a.

To znamená, že za jeden den expertní systém vytvořil zisk v průměru 358,12 USD, což představuje 1,791% vložených peněz. Uvedený zisk však nelze extrapolovat na celý rok (dostali bychom neuvěřitelné 6,5-násobné zhodnocení vložených peněz), protože jednak burza nepracuje o víkendech a svátcích, jednak ne každý burzovní den je vhodný k obchodům.

Proto jsem se pokusil o dolní realističtější vyčíslení dosaženého zisku. Zanedbám-li skutečnost, že expertní systém několik dní nemohl běžet z důvodu, že se na něm pracovalo, bylo zisku 3 939,39 USD dosaženo za 27 kalendářních dní. To znamená, že průměrný zisk na jeden kalendářní den činí asi 145,90 USD, což reprezentuje průměrné denní zhodnocení o 0,729%. Kdybychom tuto hodnotu extrapolovali na celý kalendářní rok, vyšel by nám průměrný roční zisk 266,5% p.a.

Přehledně jsou dosažené výsledky shrnuty v následující tabulce. I s vědomím, že výsledky mohou být zkreslené krátkou dobou sledování a případnými náhodnými vlivy na trhu, přesto dosažený zisk řádu 250% p.a. při relativně nízkém riziku pokládáme za mimořádně úspěšný. Jen pro srovnání připomínám, že současné úrokové sazby u České spořitelny jsou do 2% p.a. (mínus daň).

7. Další vývoj

V současné době připravuji další práce na zlepšení výsledků a bezpečnosti popsaného expertního systému. Současně s tím probíhají pokusy o implementaci ES do českého prostředí. Největším problémem se v současnosti zdá legislativa, konkrétně otázka zdanění ztrát a zisků.

Bohužel nemohu nepřipustit, že při testování systému se objevila řada chyb. Vesměš se jednalo o chyby programátorské: například nesprávně nastavené znaménko v testu způsobilo zcela katastrofické selhání (nečekaný masivní nákup akcií v okamžiku, kdy se naopak mělo prodávat). Takže celkový výsledek, vezmu-li v úvahu i selhání a chyby při neúspěšných pokusech, je spíše záporný.

S příchodem současné celosvětové finanční krize a s tím související nestabilitou na akciových trzích jsem raději další pokusy odložil.

Poděkování

Článek vznikl za podpory Rozvojového projektu VVŠ č. 941711.4/17 na rok 2009.

Literatura

- [1] Kokeš, J.: Okamžitá frekvence a Hilbert-Huangova transformace. Sborník konference Inteligentní systémy pro praxi, Lázně Bohdaneč 2006.
- [2] Kruger, CJC: Constrained Cubic Spline Interpolation for Chemical Engineering Applications. Dostupné na internetu na www.korf.co.uk/spline.pdf.
- [3] Kokeš J.: Expertní systém v reálném čase. Sborník konference Inteligentní systémy pro praxi, Seč 2004
- [4] Verner, M.: EMD – základ pro HHT. Sborník konference Inteligentní systémy pro praxi, Lázně Bohdaneč 2007.
- [5] Bursíková, S.: HHT a glykemická křivka. Sborník konference Inteligentní systémy pro praxi, Lázně Bohdaneč 2007..

Kontaktní adresa

Ústav 12110.3, Fakulta strojní, ČVUT v Praze, Technická 4, 166 07 Praha 6, Česká republika.
josef.kokes@fs.cvut.cz