KLASIFIKÁTOR IZOLOVANÝCH SLOV NA BÁZI UMĚLÉ NEURONOVÉ SÍTĚ

David Juráček

PČR MŘ Brno

Abstrakt

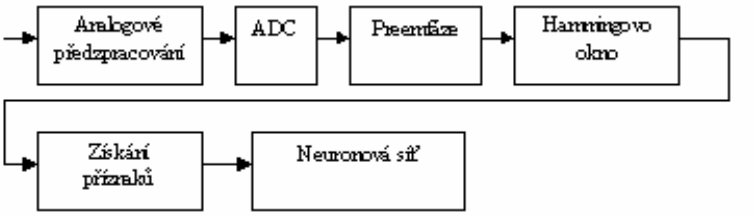
V příspěvku je demonstrováno využití umělé neuronové sítě pro klasifikaci izolovaných slov od vybrané skupiny řečníků. V programovém prostředí MATLAB byl sestaven klasifikátor izolovaných slov na bázi umělé neuronové sítě. Experimentálním způsobem byl stanoven optimální počet neuronů ve struktuře klasifikátoru. Veškeré výpočty s vyjímkou nahrání a digitalizace vzorků řeči byly provedeny v programovém prostředí MATLAB.

1 Úvod

Analýza řečových signálů [1] je v současnosti dynamicky rozvíjející se součást oblasti zpracování signálu. Vzhledem k velkému množství různých variant vyjádření stejného obsahu řeči a skutečnosti, že způsob rozpoznávání řečového signálu člověkem není v současnosti jednoznačně popsán, tvoří analýza řeči jednu z vhodných oblastí použití umělých neuronových sítí.

2 Koncepce klasifikátoru řeči

Na základě poznatků publikovaných v [2, 3, 4, 5, 6, 7] byla navržena originální koncepce klasifikátoru řeči pro rozpoznávání jednotlivých mluvčích. Klasifikátor řeči se skládá z těchto šesti bloků: (1) blok pro analogové předzpracování vzorku řeči, (2) A/D převodník, (3) blok preemfáze, (4) blok segmentace Hammingovým oknem, (5) blok pro extrakci charakteristických příznaků řeči (6) klasifikátor na bázi umělé neuronové sítě. Blokové schéma klasifikátoru řeči je uvedeno na obr. 1.



Obr. 1: Blokové schéma klasifikátoru řeči

Nyní si stručně nastíníme význam jednotlivých bloků klasifikátoru řeči.

Analogové předzpracování

Pro správnou funkci klasifikátoru je žádoucí, aby ve vzorcích řečového signálu nebyly obsaženy rušivé signály. Velikost poměru výkonu signálu k šumu má významný vliv na chybovost rozpoznávání. V rámci tohoto bloku je možné vyhodnotit úroveň šumu v signálu a následně provést korekci např. metodou spektrálního odečítání, případně zvolit metodu parametrizace odolnější vůči šumu.

Analogově digitální převodník (ADC)

Analogově digitální převodník je nutno volit s ohledem na požadavky rozpoznávání řeči. Na rozpoznávání obsahu řečových signálů postačí nižší hodnoty vzorkovacího kmitočtu (f_{vz}) a kvantovacích úrovní. Obvykle se setkáme s hodnotami f_{vz} v rozmezí (8 ÷ 12) kHz a kvantování přibližně (8 ÷ 10) bity. V případě diagnostiky řečových orgánů se požadavky na ADC zvýší na hodnoty $f_{vz} \sim 22$ kHz a kvantování až 16 bity.

Preemfáze

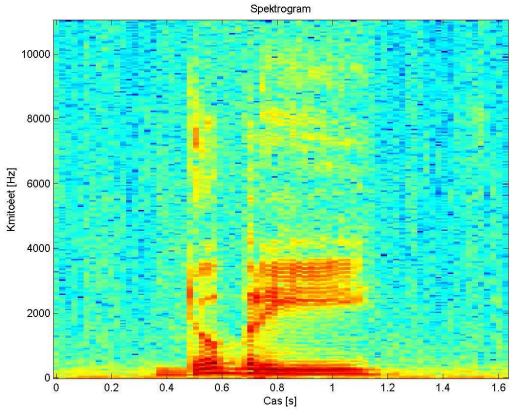
Z charakteristik spektrální výkonové hustoty řeči vyplývá poznatek, že tato křivka dosahuje maxim v okolí 300 Hz pro mužský hlas a přibližně v 500 Hz pro hlas ženský. V úseku v rozmezí 500 až 2000 Hz se nachází převážná část energie podstatná pro rozpoznání řeči. Vzhledem k poklesu úrovně spektrální výkonové hustoty řečového signálu, zaznamenané od kmitočtu 500 Hz se sklonem přibližně 6 dB na oktávu, se tato disproporce v systémech pro zpracování řeči odstraňuje zařazením korekčního článku typu preemfáze.

Hammingovo okno

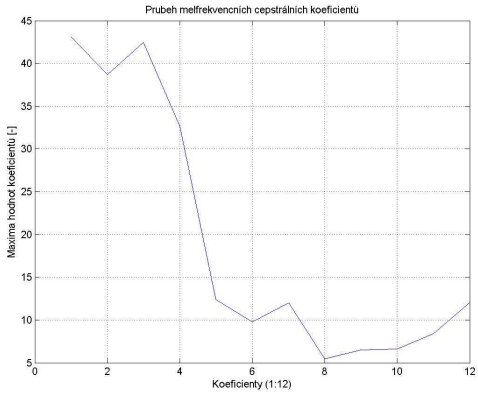
Pro analýzu řečových signálů se jako nejvýhodnější funkce pro účely segmentace jeví tzv. Hammingovo okno [4]. Volbou okna s výrazným potlačením postranních laloků v modulu kmitočtového spektra váhovací funkce dosáhneme lepší potlačení periodicky se opakujících složek analyzovaného signálu. Optimálních výsledků lze dosáhnout vhodnou volbou "překryvu" oken.

Extrakce charakteristických příznaků řeči

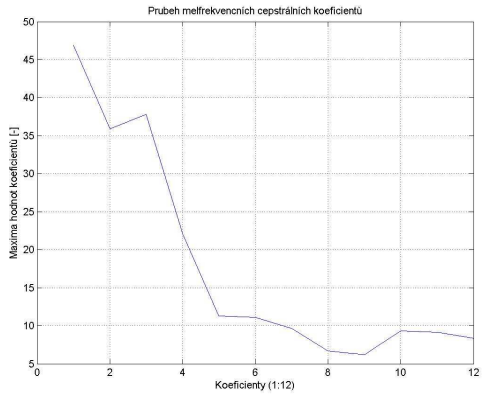
Na základě rozboru a analýzy typických charakteristik řeči a na základě poznatků uvedených v [4, 5] byla pro extrakci charakteristických příznaků řeči zvolena melfrekvenční cepstrální analýza. V programovém prostředí Matlab byl sestaven algoritmus pro výpočet 12-ti melfrekvenčních cepstrálních koeficientů (MFCC). Příklad průběhu MFCC pro izolované slovo je uveden na obr. 2.



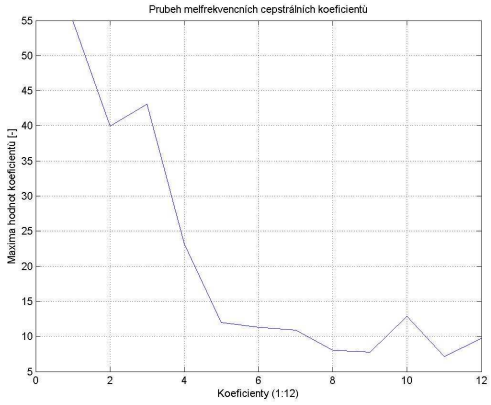
Spektrogram slova „dobrý" řečníka č. 3



MFCC slova „dobrý" řečníka č. 3



MFCC slova „dobrý" řečníka č. 2



MFCC slova „dobrý" řečníka č. 1

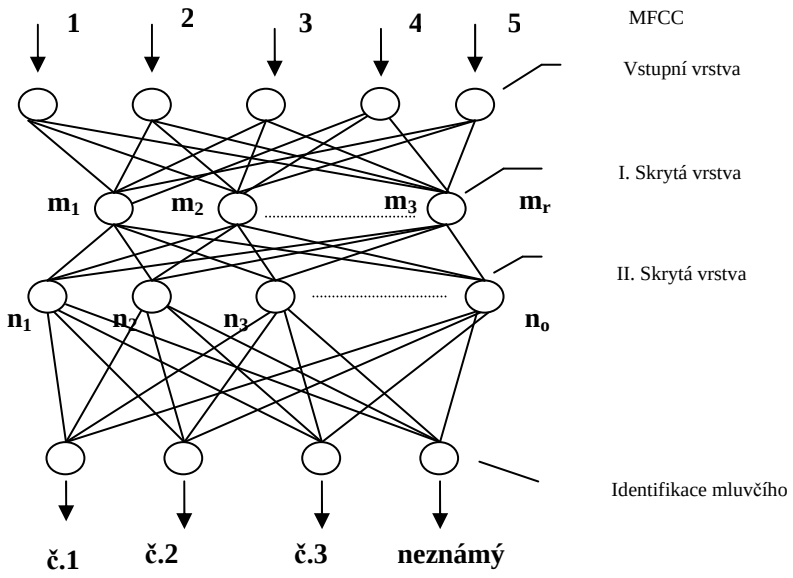
Obr. 2: Průběh melfrekvenčních cepstrálních koeficientů a spektrogramu klíčového slova „dobrý" pro tři řečníky

Na zobrazených grafech je zajímavé povšimnout si u stejného izolovaného slova (od různých řečníků) velmi podobných hodnot MFCC. Tato zobrazení potvrdila předpoklad, že zvolená metoda analýzy řeči dosahuje relativně dobrých výsledků v rozpoznávání obsahu řeči, neboť částečně odstraňuje závislost na individuálním charakteru řečníka.

Klasifikátor na bázi umělé neuronové sítě

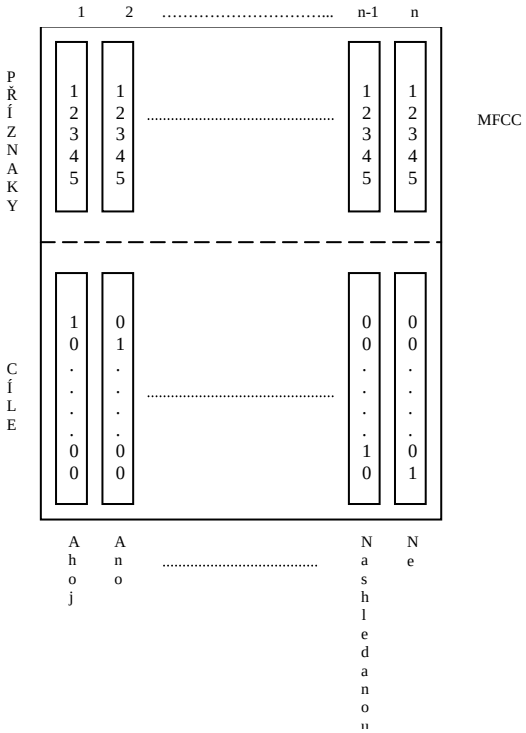
Pro účely klasifikace řeči je možno zvolit různé topologie neuronových sítí. Dopředné neuronové sítě typu backpropagation jsou používány zejména v aplikacích predikce a klasifikace [4, 6, 7]. Dle doporučení uvedených v [4, 5, 6, 7] byl sestaven klasifikátor izolovaných slov, který je tvořen umělou neuronovou sítí typu backpropagation. Umělá neuronová síť je tvořena čtyřmi vrstvami neuronů. Vstupní vrstva má pět neuronů, které odpovídají počtu pro klasifikaci zvoleným MFCC, v první skryté vrstvě má neuronová síť osm neuronů, deset neuronů se nachází v druhé skryté vrstvě a výstupní vrstvu tvoří počet neuronů odpovídající počtu klasifikovaných klíčových slov. Topologie neuronové sítě pro klasifikátor izolovaných slov je uvedena na obr. 3.

Ve vstupní vrstvě je použito pět konkrétních hodnot MFCC, výstupní vrstvu tvoří osm neuronů, což odpovídá počtu klasifikovaných slov (osmý je neznámý).



Obr. 3: Struktura dopředné umělé neuronové sítě pro identifikaci izolovaných slov

Pro strukturu umělých neuronové sítě z obr. 3 byla vytvořena trénovací množina. Trénovací množina se skládá z matice cílů (izolovaná slova) a matice příznaků (zvolené MFCC). Příklad struktury trénovací množiny je naznačen na obr. 4.



Obr. 4: Příklad struktury trénovací množiny pro izolovaná slova

3 Výsledky experimentu

Pro experimentální ověření činnosti klasifikátoru izolovaných slov na bázi umělé neuronové sítě byla zvolena nejčastěji vyskytující se slova v běžné české mluvě (ahoj, ano, dobrý, den, dobrý den, nashledanou, ne). Byla provedena klasifikace těchto sedmi izolovaných slov od sedmi mluvčích, z nichž byly čtyři vzorky řeči od mužů a tři od žen. Výsledky klasifikace izolovaných slov jsou uvedeny v tab. 1.

Tab. 1: Výsledky klasifikace izolovaných slov od sedmi mluvčích

Slovo	*Ahoj*	*Ano*	*Dobrý*	*Den*	*Dobrý den*	*Nashledanou*	*Ne*	*Celkem*
úspěšnost	*62,46 %*	*37.54 %*	*42,15 %*	*46,3 %*	*45,38 %*	*45,85 %*	*36,46 %*	*45,16 %*

Dále byla experimentálně sledována úspěšnost klasifikace izolovaných slov pro jednotlivé mluvčí. Výsledky experimentu jsou uvedeny v tab. 2.

Tab. 2: Úspěšnost klasifikace izolovaných slov pro jednotlivé mluvčí

Řečník	*Muž 1*	*Muž 2*	*Muž 3*	*Muž 4*	*Žena 1*	*Žena 2*	*Žena 3*	*Celkem*
úspěšnost	*67,14 %*	*99,71 %*	*74,57 %*	*57,71 %*	*87,71 %*	*67,14 %*	*58,86 %*	*73,27 %*

Z výsledků experimentu vyplývá, že v případě klasifikace klíčových slov bylo dosaženo větší úspěšnosti než v případě rozpoznávání mluvčího s využitím klíčových slov. Tato skutečnost byla před provedením pokusů samotných předpokládána, a to zejména ze dvou důvodů.

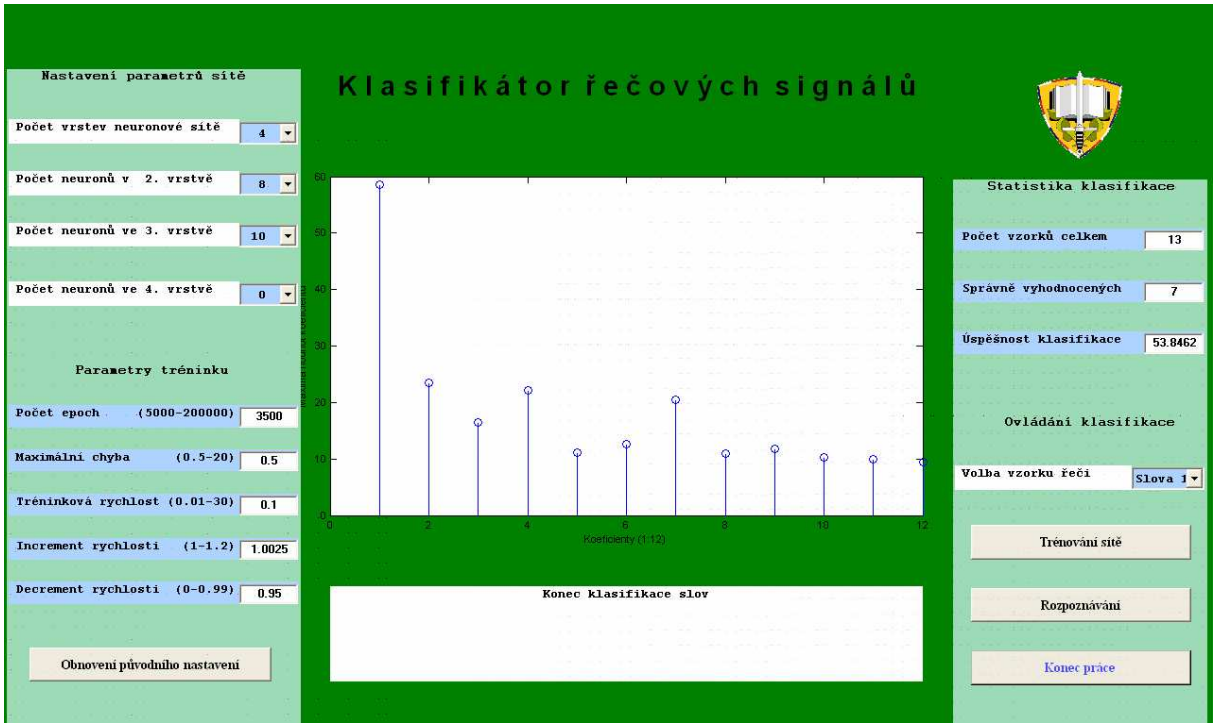
Důvodem prvním je systém skladby trénovací množiny. Při klasifikaci klíčových slov bylo zpočátku dosaženo výsledků o přibližně (15 ÷ 20) % horších, než v prezentovaném případě. Zlepšení bylo dosaženo změnou systému skladby trénovací množiny.

Dalším důvodem je, že melfrekvenční cepstrální analýza do značné míry odstraňuje závislost vlastního obsahu řeči na individuálním charakteru projevu, určitá “zbytková” závislost však zůstane zachována. Nižší percentuální hodnoty dosažených výsledků této skutečnosti nasvědčují.

Ve vstupní vrstvě je použito pět konkrétních hodnot MFCC, výstupní vrstvu tvoří osm neuronů, což odpovídá počtu klasifikovaných slov.

4 Demonstrace činnosti klasifikátoru v programovém prostředí Matlab

Aby bylo možné demonstrovat činnost klasifikátoru izolovaných slov na bázi neuronové sítě, bylo navrženo programové rozhraní v programovém prostředí Matlab (obr. 5). Toto programové rozhraní bylo vytvořeno v rámci řešení projektu [8]. Navržené rozhraní je využíváno pro účely výuky studentů na Univerzitě obrany.



Obr. 5 Programové rozhraní pro klasifikaci izolovaných slov

V levé horní části zobrazeného okna je možno volit parametry sítě a tréninku. Naprogramovány jsou varianty tří, čtyř a pětivrstvých neuronových sítí (s jednou, dvěma, třemi skrytými vrstvami). Popup menu v levé horní části volíme počet vrstev sítě. Nižší tři popup menu slouží k volbě počtu neuronů ve skrytých vrstvách sítě. Pod těmito volbami se nachází pětice editačních oken, kde je možno zadat (měnit) parametry tréninku umělé neuronové sítě. V závorkách jsou uvedeny orientační meze volby, ve kterých má trénink sítě smysl nastavovat. V levé dolní části okna se nachází tlačítko obnovení původního nastavení, které síti vrací přednastavené (odladěné) parametry tréninku.

V pravé střední části okna je popup menu, kterým lze vybírat požadovanou činnost sítě (klasifikaci fonémů – volba Muž 1 až Žena 3, nebo rozpoznávání mluvčích – volba Slova 1 až Slova 7).

Pod tímto menu se nachází tlačítka Trénink sítě a Rozpoznávání. Po výběru požadované činnosti sítě je nutno síť natrénovat. Trénink sítě se aktivuje stejnojmenným tlačítkem. Po jeho spuštění se objeví okno, které zobrazuje hodnotu součtu čtverců kvadratické chyby a velikost kroku učení sítě v závislosti na počtu vykonaných epoch. Po dosažení zadané hodnoty maximální chyby (příp. po vykonání nastaveného počtu epoch v případě, že nastavená chyba je tak malá, že jí nelze dosáhnout) se v dolní střední části zobrazí hlášení o úspěšnosti tréninku sítě (síť pracuje/nepracuje odpovídajícím způsobem).

Nyní je možno přistoupit k vlastní klasifikaci. Popup menu Volba vzorku řeči vybereme vzorek, který chceme klasifikovat a tlačítkem Rozpoznávání tento proces spustíme.

V průběhu rozpoznávání se v grafu uprostřed okna zobrazují hodnoty melfrekvenčních cepstrálních koeficientů (všech dvanácti, do sítě jsou přivedeny pouze koeficienty s č. 2, 5, 7, 9, 10), které jako parametry přichází do sítě. Po dokončení klasifikace se v okně pod grafem zobrazí hlášení o ukončení procedury a v pravém horním rohu okna v sekci Statistika klasifikace se zobrazí údaje o úspěšnosti klasifikace. Rozhraní se uzavírá tlačítkem Konec práce.

5 Závěr

Experimentální výsledky ukázaly, že klasifikátor řeči založený na bázi dopředné umělé neuronové sítě typu backpropagation lze realizovat. Jeho využití je možné zejména v případech, kdy počet osob, jejichž vzorky budeme analyzovat, bude velmi nízký.

S využitím rozvíjejících se metod parametrizace řečových signálů lze předpokládat snížení závislostí individuálního charakteru řečníka na obsah řeči na hodnoty, na které se zvládne umělá neuronová síť dynamicky adaptovat.

Vytvořené programové rozhraní dovoluje uživateli demonstrovat činnost umělých neuronových sítí v programovém prostředí Matlab a seznamuje uživatele s charakteristikami reálných vzorků řeči.

Literatura

[1] J. Černocký. Temporal processing for feature extraction in speech recognition. VUT, Brno, 2003.
[2] L. Rabiner,B. H. Juang. Fundamentals of speech recognition. Prentice Hall, London, 1993.
[3] D. P. Morgan, Ch. L. Scofield. Neural Networks and Speech Processing. Kluwer Academic Publishers, Boston, 1991.
[4] Y. Bengio. Neural Networks for Speech and Sequence Recognition. International Thomson Publishing, London, 1996.
[5] M. G. Rahim. Artificial Neural Networks for Speech Analysis/Synthesis. Chapman and Hall, London, 1994.
[6] M. Richterova. Signal Modulation Recognizer Based on Method of Artificial Neural Networks. In Proceeding of Symposium PIERS, Hangzhou, China, 2005.
[7] M. Richterová, A. Mazálek, K. Pelikán. Modulation Classifier of Digitally Modulated Signals Based on Method of Artificial Neural Networks. In Proceeding of the WSEAS Conferences: 4th WSEAS Int. Conf. on AEE'05. Prague, 2005.
[8] D. Juráček. Klasifikátor řeči. Diplomová práce. Univerzita obrany, Brno, 2005.
[9] H. Demuth, M. Beale. Neural Network Toolbox. For Use with MATLAB. User's Guid., ver. 4. The MathWorks, Inc., MA 01760-2098.

Kontakt
Ing. David Juráček
PČR MŘ Brno, Cejl 4/6, 602 00 Brno
E-mail: Davidjk@seznam.cz