UKÁZKA ROZPOZNÁVÁNÍ ŘEČI S VYUŽITÍM KLASIFIKACE NEURONOVOU SÍTÍ

Vladimír Hlaváč

ČVUT v Praze, Fakulta strojní, hlavac@fs.cvut.cz

Abstrakt: Článek demonstruje použití klasifikace neuronovou sítí na data, získaná FFT transformací a pásmovým rozdělením do složek vektoru ze zvukového záznamu několika samohlásek, zaznamenaných zvukovou kartou. Experiment byl proveden za účelem porovnání s metodou SVM pro klasifikaci dat.
Klíčová slova: Rozpoznávání zvuku, FFT, Neuronová síť, Klasifikace dat.

1. Úvod

Rozpoznávání mluveného slova pomocí analýzy frekvenčního spektra bylo zkoumáno již od sedmdesátých let, dlouho před rozšířením počítačů. Použití výpočetních metod popisuje již například profesor Josef Psutka v roce 1995 [1]. Jednou ze základních je použití rychlé Fourierovy transformace a nerovnoměrné (logaritmické) rozdělení pásem, kdy při sečtení amplitud v těchto pásmech vznikne vektor, který je pro danou hlásku charakteristický. Tímto způsobem lze s minimální chybovostí rozlišit přinejmenším vyslovované samohlásky za podmínky jednoho mluvčího (pro souhlásky a více mluvčích je nutné použít další metody, popsané v [1]).

Tento článek popisuje, jak získat tyto vektory, a demonstruje jejich rozpoznávání neuronovou sítí. Byl napsán jako podklad pro další práce, kdy studenti řeší obdobný problém metodou SVM, za účelem jejich porovnání.

2. Získávání dat pro rozpoznávání

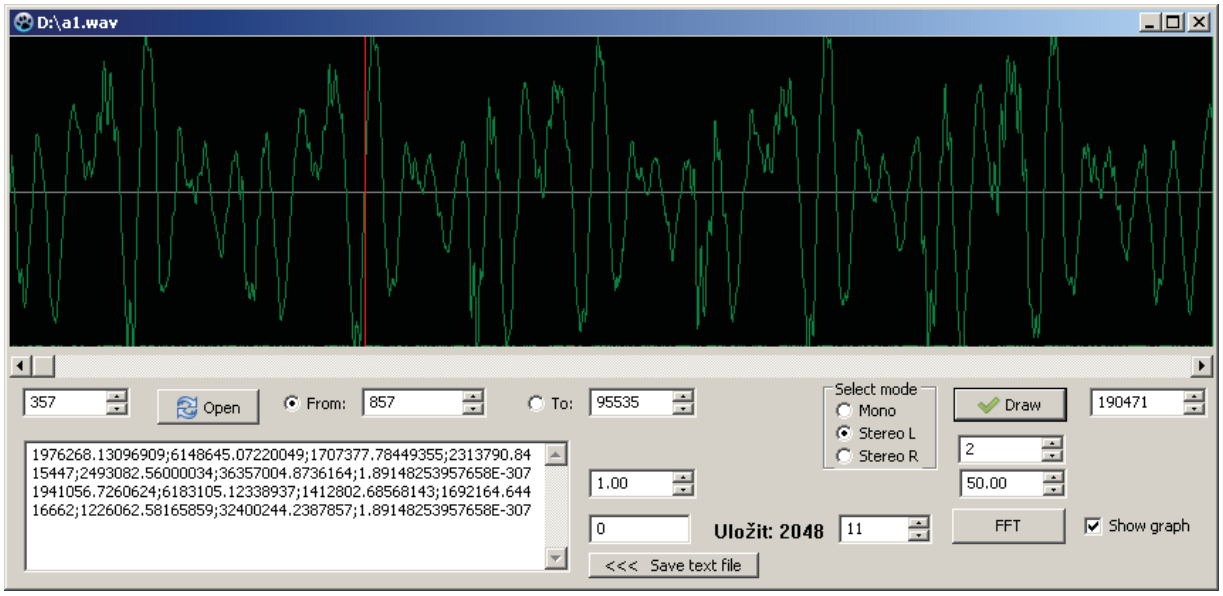
2.1 Data v časové oblasti

Data jsou načtena pomocí zvukové karty a programu Záznam zvuku, který je součástí Windows. Data byla zaznamenána ve formátu wav, který není pakovaný. Pro záznam byly vybrány hlásky, které lze vyslovovat v délce sekund, po jisté subjektivní úvaze samohlásky a, e, i, o, u a souhláska s. Záznam neobsahuje začátek ani konec vyslovování.

2.2 Převod do frekvenční oblasti a rozdělení do pásem

Funkce FFT [2] by byla nejrychlější v Matlabu, ale je třeba nejdříve načíst data, a stejně jako v jazyce Python, musela by být použita řada knihoven. Protože se nejedná o prestižní projekt, bylo použito jednodušší řešení a vše bylo napsáno v Pascalu, kde je k dispozici výkonné RAD prostředí Lazarus. Ukázkové řešení FFT pro Delphi lze nalézt na internetu, ale bylo použito z archivu autora.

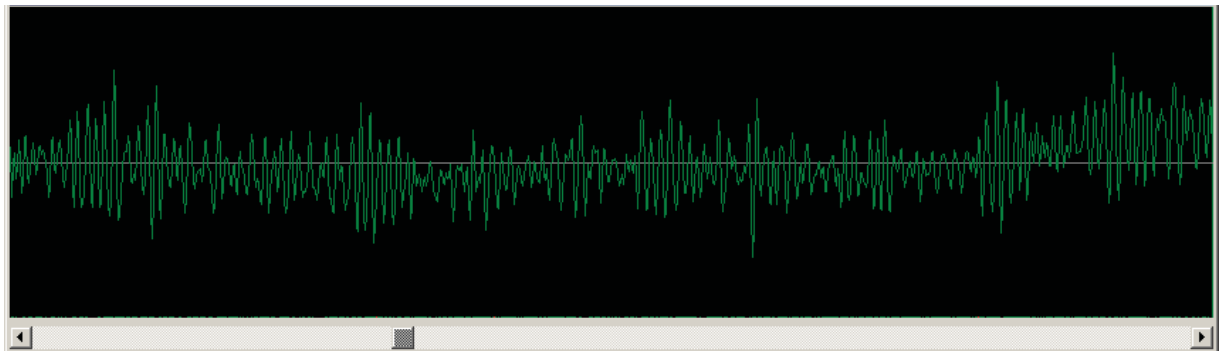
Program umožňuje načíst přímo soubory formátu wav v mono nebo stereo verzi pro 16bitové vzorky. Po načtení je třeba nastavit, zda se jedná o záznam mono nebo stereo a v druhém případě, který kanál číst. Všechny analýzy byly provedeny pro stereo záznam a levý kanál.



Obr. 1 Program pro výpočet FFT, načtena hláska "a". Po kliknutí na graf se doplní orientační červená čára, v daném bodě se zahájí čtení dat pro FFT (počet vzorků je dán nastavením exponentu základu „2“ v okně nalevo od tlačítka FFT, zde 11 – odpovídá 2048, ale používalo se 4096), výsledné frekvence se po intervalech sečtou (viz text) a výsledek se zapíše do okna textového editoru vlevo dole. Odtud je lze kopírovat přes schránku, nebo tlačítkem všechna najednou zaznamenat do souboru. Výměnou zpracovávaných dat (zelený graf) se toto okno nemaže, a lze tak snadno vybírat tutéž hlásku z různých záznamů zvuku.

Po načtení dat je třeba zvolit parametry transformace. Pro dále vyhodnocované vzorky bylo použito okno o šířce 4096 vzorků (počet iterací FFT je třeba přenastavit z 11 na 12). Data na obr. 1 jsou 50× zmenšená a vynesena jako 2 vzorky na pixel.

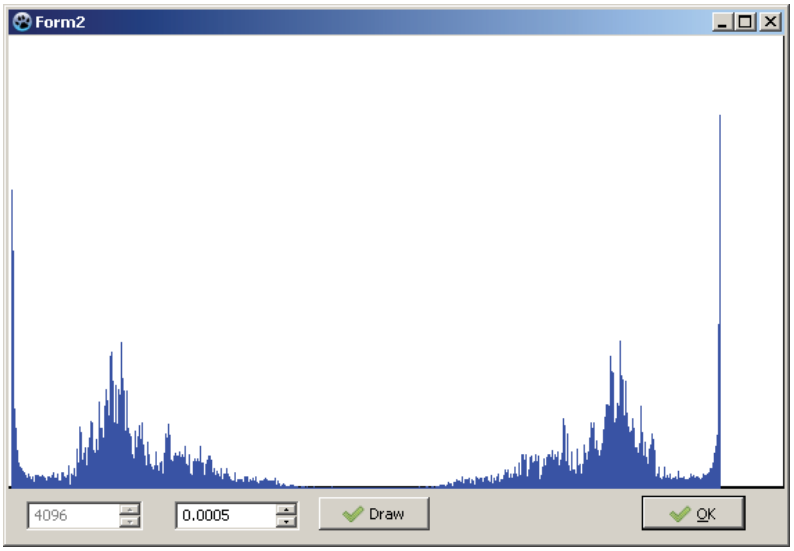
Pro nízké frekvence FFT má velký vliv poloha okna. I když funkce je zjevně periodická, FFT ve skutečnosti vyhodnocuje spektrum z dat, která vzniknou tak, že načtených (zde) 4096 vzorků se periodicky opakuje, což je velmi ovlivněno místem, kde se přesně začnou data přebírat. Tomuto efektu se nechá zamezit aplikací Hammingova okna. Ve zde popisovaném případě byl ale použit manuální postup, kdy vzorky jsou vždy načítány od místa průchodu nulou, bezprostředně předcházející lokálnímu maximu funkce (na obr. 1 červená čára v grafu, označí se kliknutím myši). Oba postupy vedou k obdobným výsledkům.



Obr. 2 Záznam hlásky "s"

Pro samohlásky „a“ a „e“ vychází převaha nízkých frekvencí, pro „o“, „u“ a „i“ jsou pak vyšší. Všeobecně se vyšší frekvence objevují hlavně u souhlásek, ale zde jsou stabilní průběhy jen u „s“, „š“, „z“ a „ž“. Některé souhlásky mají hodně krátký průběh („k“), u jiných se frekvence v průběhu hlásky mění („l“).

Po stisknutí tlačítka FFT je zobrazena funkce (obr. 3). Samotný vektor je do okna výsledků zapsán po kliknutí do grafu záznamu v časové oblasti, aby šlo rychle vybrat dostatečný počet dat.



Obr. 3 Průběh spektra hlásky "s" (levá polovina grafu). Výpočtem rychlé Fourierovy transformace vyjdou komplexní hodnoty, které pro každou dílčí frekvenci reprezentují vektor otočený podle fázového zpoždění dílčí frekvence oproti začátku skenovaného intervalu. FFT z reálných hodnot je vždy osově symetrická, kromě první hodnoty, která má vazbu na nevyváženost stejnosměrné hodnoty signálu (program ji v dalším ignoruje). Zobrazeny jsou ale velikosti dílčích frekvencí (magnitudy), získané jako odmocnina ze součtu druhých mocnin reálné a imaginární hodnoty.

Vektor (jednorozměrná matice) hodnot se ze spektra získá tak, že do první složky se sečte prvních 10 hodnot, do další následujících 20 (frekvence č. 11 až 30), pak dalších 40, atd. Způsob rozdělení na frekvenční pásma má velký vliv na výsledné rozpoznávání.

Program umožňuje rychle zaznamenat velké množství vzorků a následně je uložit do textového souboru. Byl uložen různý počet vzorků, od 63 pro „e“ po 140 pro „i“. Jednotlivé textové soubory byly předzpracovány programem MS Excel. Poslední prvek vektoru, který obsahoval pravou polovinu spektra, byl odstraněn, a ostatní složky byly poděleny takovým číslem, aby součet byl 10000. Takto upravená data se hodí na porovnání různých způsobů rozpoznávání.

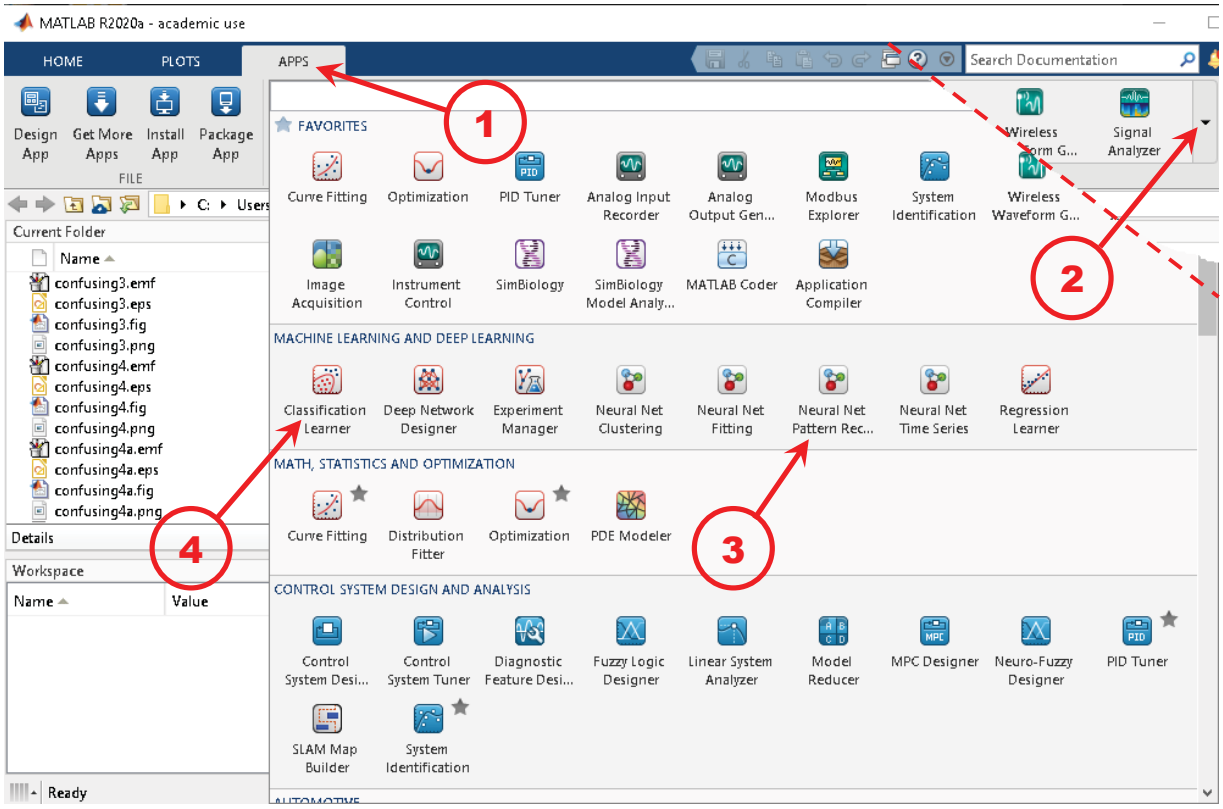
Pro zpracování v Matlabu byly doplněny sloupečky s příslušností dat (pro daný vzorek se doplní do řádku nuly kromě místa, které odpovídá dané hlásce, kde se doplní jednička), všechny tabulky spojeny a zaznamenány do textového souboru. Pro načtení Matlabem řádky nemají záhlaví.

3. Rozpoznávání neuronovou sítí

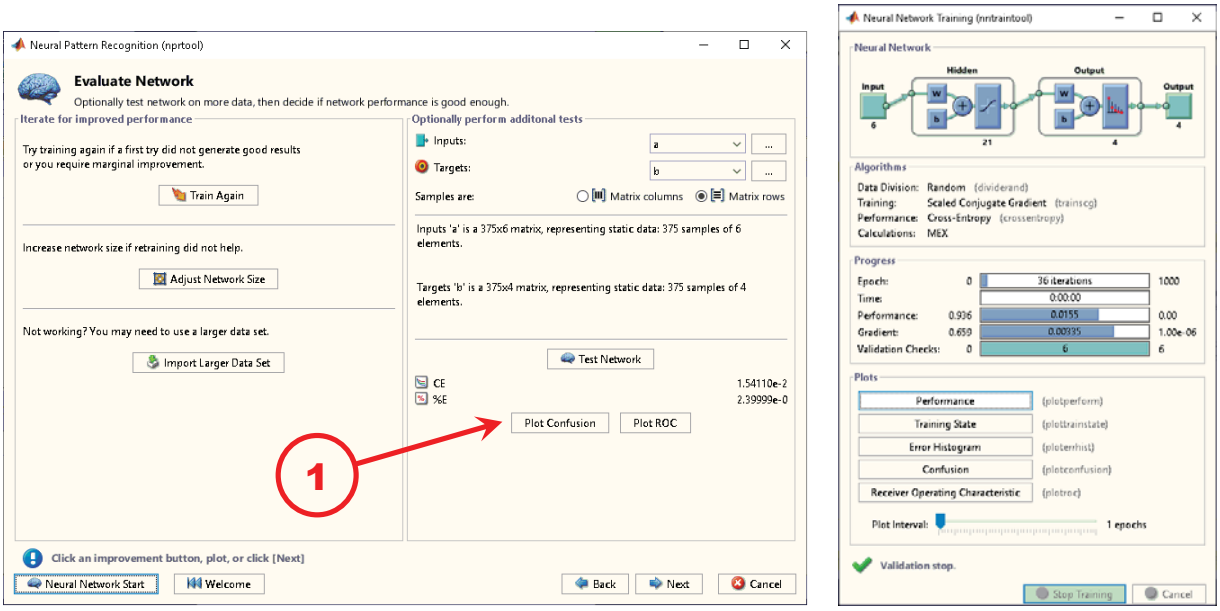
Aplikace pro pokročilé zpracování dat lze v Matlabu nalézt pod záložkou APPS. Seznam aplikací otevřeme a vybereme Neural Net Pattern Recognition, viz obr. 4. Také je možné ji přímo spustit příkazem nprtool.

Aplikace nprtool připomíná vzhledem aplikaci nftool. V několika krocích zvolíme zdrojová data, jejich rozdělení na trénovací, validační a testovací sadu, metodu trénování, počet neuronů ve skryté vrstvě a můžeme zkusit první pokus o trénování. Po natrénování aplikace umožňuje přenastavit parametry nebo změnit data a zkusit trénování opakovat (to lze i beze změn, metoda je stochastická a získáme jiné výsledky). Na závěr je možné výslednou síť vyexportovat, například jako funkci v Matlabu, a tu pak používat pro klasifikaci dalších dat.

Data (popis vyhodnocovaných dat viz předchozí kapitola) byla tvořena šesti sloupci součtu magnitud frekvencí v jednotlivých pásmech, ke kterým bylo doplněno šest sloupců správné klasifikace. Nejprve byla vyhodnocována data pouze pro čtyři písmena, „a-e-i-s“, kdy byly samozřejmě nadbytečné sloupce vynechány. Tato zkrácená tabulka měla 375 vzorků. Pro trénování byly rozděleny mezi trénovací, validační a testovací sadu v poměru 80% – 10% – 10%, tedy 299 – 38 – 38 (default je 70% – 15% – 15%). Bylo nastaveno 21 neuronů ve skryté vrstvě. Matlab provedl 26 iterací. Výsledky trénování jsou na obr. 6, obr. 7 a obr. 8.



Obr. 4 Vyhledání aplikace pro klasifikaci pomocí Neuronové sítě. Kliknutím na záložku APPS (1) se přepneme na kartu aplikací. Vpravo (2) otevřeme nabídku (čárkovaná čára odděluje vzhled okne před kliknutím, vpravo nahoře, a po rozevření nabídky, většina obrázku). Většina nabídky není zobrazena (posuvník vpravo). Kroužek (3) označuje použitý modul pro klasifikaci neuronovou sítí. Kroužek (4) pak aplikaci pro metodu podpůrných vektorů, SVM, příkaz Matlabu classificationLearner.

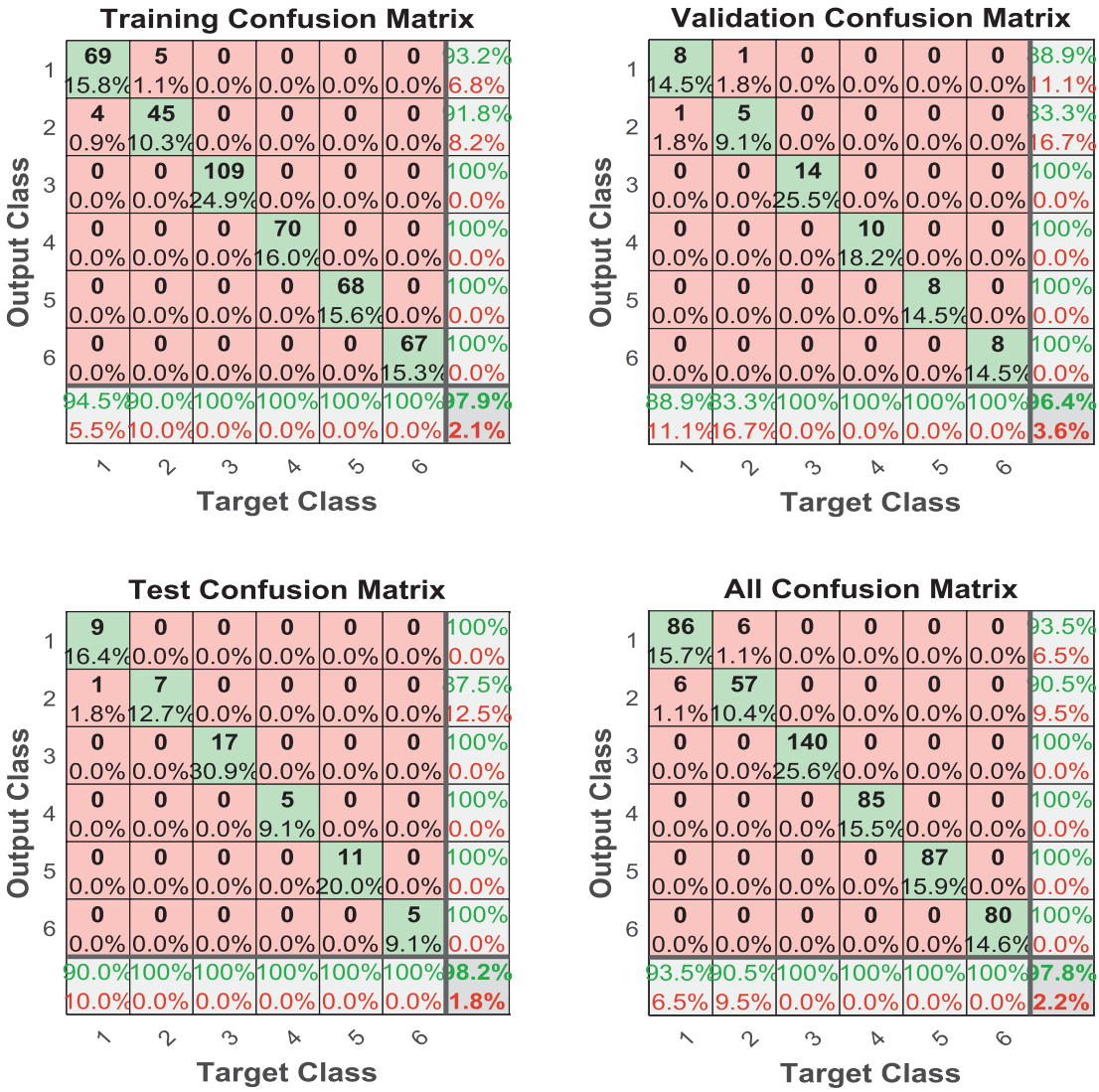


Obr. 5 Předposlední okno aplikace nprtool. Stisknutím označeného tlačítka Plot Confusion získáme tabulky, které jsou na obr. 6. (vpravo okno z průběhu trénování sítě; byla použita metoda Scaled Conjugate Gradient).

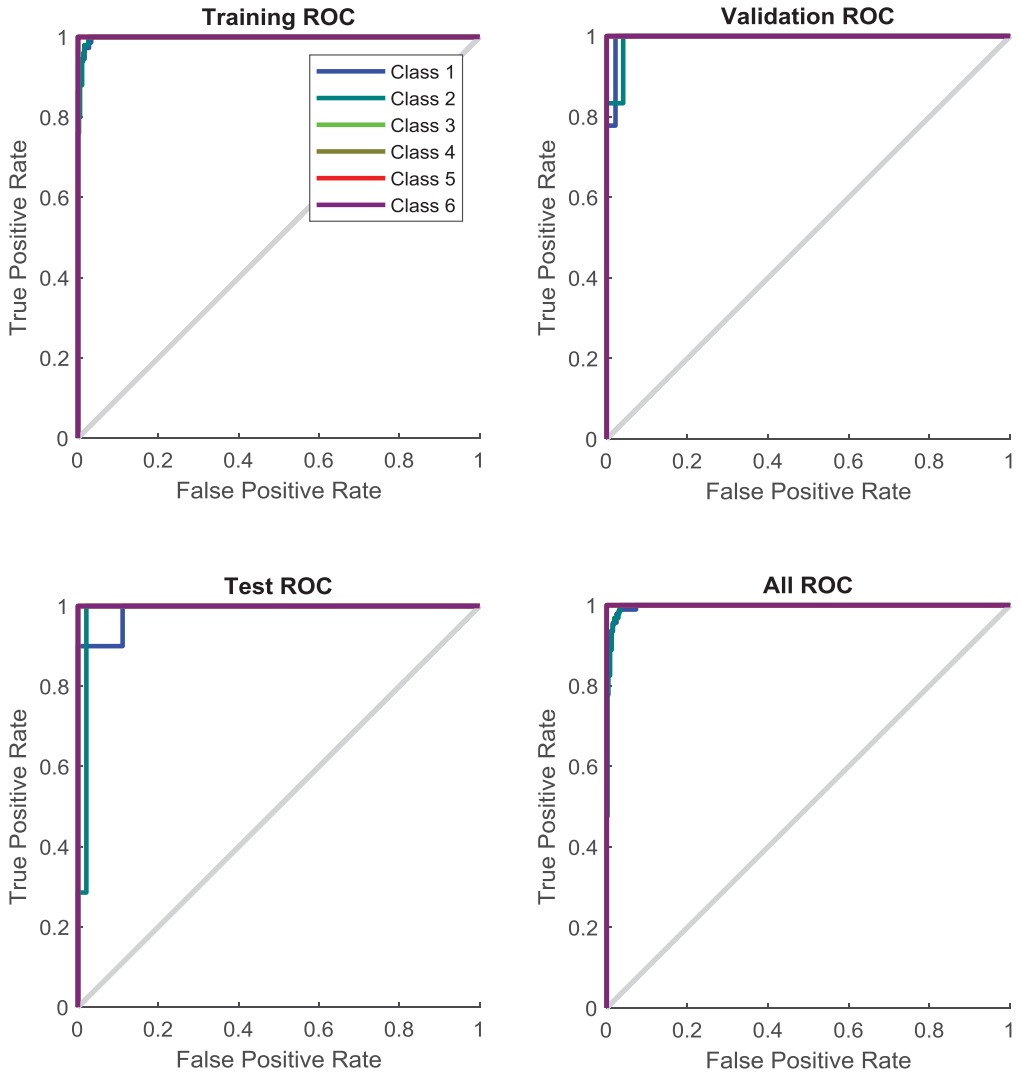
Obr. 6 Výsledky trénování. Svisle výsledek klasifikace, vodorovně správná hodnota. Dochází jen k záměnám hlásek „a“ a „e“, což může být poměrně hrubým rozdělením začátku frekvencí.

Obr. 7 Po sloučení hlásek "a" a "e" do jedné skupiny máme naprosto bezchybné rozpoznávání.

Obr. 8 Původní data, původní nastavení Matlabu (263 trénovacích, 56 validačních i testovacích, 10 neuronů ve skryté vrstvě, Matlab provedl 20 iterací). Výsledek je prakticky stejný, resp. rozdíl může být dán stochastickým charakterem metody.



Obr. 9 Úplná sada dat (1=a, 2=e, 3=i, 4=o, 5=u, 6=s). Výsledky jsou obdobné. Celý soubor má 576 řádek, ty se dělily na 437 trénovacích, 55 validačních a stejný počet testovacích. Skrytá vrstva byla nastavena na 21 neuronů. Proběhlo 46 iterací. I v tomto případě nejsou rozpoznány hlásky „a“ a „e“, zatímco ostatní jsou bezchybně rozlišeny.



Obr. 10 Plot recognition, vynesení přiřazení čárově. Tatáž data jako na obr. 9. Tento druh zobrazení má smysl, pokud je příliš mnoho tříd objektů na rozpoznávání.

4. Závěr

Byla připravena data pro testování modulu nprtool, která by se měla využít i pro metodu podpůrných vektorů (SVM), v Matlabu funkce fitecoc (error-correcting output codes, ECOC). Data se ukázala po drobné úpravě (sloučení hlásek „a“ a „e“) plně rozpoznatelná neuronovou sítí, měla by tedy být rozpoznatelná i pomocí SVM, pravděpodobně i ve zjednodušené lineární formě.

5. Reference

[1] Psutka, Josef: Komunikace s počítačem mluvenou řečí. Academia, Praha, 1995. ISBN 80-200-0203-0

[2] Erickson, Jeff: Algorithms. Amazon, 2019. ISBN 1792644833

Selected article from
Tento dokument byl publikován ve sborníku

Nové metody a postupy v oblasti přístrojové techniky, automatického řízení a informatiky 2020

New Methods and Practices in the Instrumentation, Automatic Control and Informatics 2020

14. 9. – 16. 9. 2020, Zámek Lobeč

ISBN 978-80-01-06776-5

Web page of the original document:
http://iat.fs.cvut.cz/nmp/2020.pdf

Obsah čísla/individual articles:
http://iat.fs.cvut.cz/nmp/2020/

Ústav přístrojové a řídicí techniky, FS ČVUT v Praze, Technická 4, Praha 6