

Neuronové sítě
pro humanitní studia

Vlasta Radová
Západočeská univerzita v Plzni
katedra kybernetiky

Neuronové sítě pro humanitní studia – rozpoznávání mluvené řeči

Komunikace s počítačem mluvenou řečí

- n Účelem každé komunikace je předání (resp. výměna) požadovaných informací mezi komunikujícími subjekty.
- n Počítač umí efektivně pracovat s informací uloženou ve formě textu. Abychom tedy umožnili komunikaci mezi člověkem a počítačem, musíme v první řadě vyřešit 2 základní úlohy:
 - n převod řeči na text – **rozpoznávání mluvené řeči**
 - n převod textu na řeč – **počítačová syntéza řeči**

6.3.2009 Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky 2

Neuronové sítě pro humanitní studia – rozpoznávání mluvené řeči

Klasifikace (rozpoznávání) předmětů a jevů

- n Schéma procesu klasifikace (rozpoznávání) **řečových signálů**

```

  graph LR
    A["předmět  
vyslovené slovo"] --> B["vytvoření  
formálního  
popisu"]
    B -- "reprezentace  
předmětu  
(obraz)" --> C["klasifikátor"]
    C -- "informace  
o třídě" --> D["napsané  
slovo"]
  
```

6.3.2009 Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky 3

Neuronové sítě pro humanitní studia – rozpoznávání mluvené řeči

Problém volby příznaků

- n Zatím neexistuje obecná metoda, která by dokázala určit jaké příznaky na daném předmětu měřit. Teoreticky je možné na předmětu měřit „všechno, co jde“, takový postup je však nereálný (komplikace s technikou realizací klasifikátoru).

↓

Příznaky se vždy musí volit na základě konzultace s odborníkem z dané problemové oblasti.

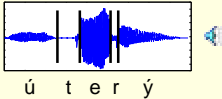
Pro výběr vhodných příznaků pro rozpoznávání mluvené řeči je třeba rozumět tomu, co je to řečový signál.

6.3.2009 Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky 4

Neuronové sítě pro humanitní studia – rozpoznávání mluvené řeči

Co je to řečový signál

Příklad – slovo „úterý“



- n v **textové formě** lze slovo „úterý“ vyjádřit pomocí 5 čísel (kódů z ASCII tabulky):
250 116 101 114 253
- n **akustickou formu** slova je třeba uložit jako posloupnost několika tisíc čísel:
-8 -4 6 -2 9 0 -4 -1 -3 -1 1 0 -17 -23 -30
-35 -45 -48 -49 -42 -39 -27 0 13 25 42
66 63 59 70 66 60 56 41 25 18 ...

6.3.2009 Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky 5

Neuronové sítě pro humanitní studia – rozpoznávání mluvené řeči

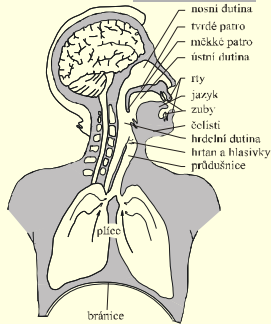
Co je to řečový signál

- n Z pohledu přenosu slovní informace je akustický řečový signál značně redundantní, protože obsahuje informaci o
 - n vyřčených slovech,
 - n výšce a barvě hlasu,
 - n tempu řeči,
 - n intonaci řeči
- a „surová“ reprezentace akustického řečového signálu tak „zabírá příliš mnoho místa“.
- n Aby bylo možné počet příznaků potřebných pro reprezentaci řečového signálu snížit, je třeba rozumět procesu, jakým je lidská řeč vytvářena.

6.3.2009 Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky 6

Proces vytváření řeči

- n Zdrojem řečových kmitů, které jsou fyzikální reprezentací řeči, jsou lidské řečové orgány – hlasivky, dutina hrdelní, nosní a ústní, měkké a tvrdé patro, zuby, jazyk a rty. K těmto orgánům je však nutno ještě připočítat základní zdroj hlasové energie, tj. plíce a s nimi funkčně spjaté dýchací svaly.



6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky

7

Proces vytváření řeči (pokračování)

- n Zdrojem **znělých zvuků** jsou kmitající hlasivky.
- n Frekvence kmitů hlasivek závisí na délce, síle a svalovém napětí hlasivek a určuje **základní tón lidského hlasu**.
- n Pro většinu lidí se základní hlasivkový tón pohybuje v rozmezí 80 až 300 Hz, může se ale měnit v rozsahu až 33 - 3100 Hz. U žen je v průměru 2x vyšší než u mužů.

6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky

8

Proces vytváření řeči (pokračování)

- n Kromě základního hlasivkového tónu se v akustickém spektru samohlásek objevuje řada vyšších zesílených tónů, které vznikají rezonancí v dutinách hlasového traktu.
- n Tyto tóny se nazývají **formanty**.
- n Frekvence formantů závisí především na velikosti a tvaru dutiny ústní.

6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky

9

Proces vytváření řeči (pokračování)

Hodnoty prvních tří formantů pro české samohlásky

samohláska	F_1 [Hz]	F_2 [Hz]	F_3 [Hz]
u	300 – 500	600 – 1000	2400 – 2900
o	500 – 700	900 – 1200	2500 – 3000
a	750 – 1100	1100 – 1500	2500 – 3000
e	500 – 700	1500 – 2000	2500 – 3000
i	300 – 500	2000 – 3000	2600 – 3000

6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky

10

Proces vytváření řeči (pokračování)

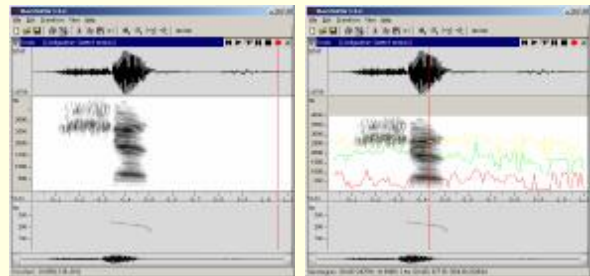
- n **Souhlásky** jsou vytvářeny vzduchovou turbulencí, která vzniká třením výdechového proudu vzduchu z plic o překážku vytvořenou artikulačními orgány a projevuje se přítomností charakteristického šumu v akustickém spektru hlásek.

6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky

11

Reprezentace řeči

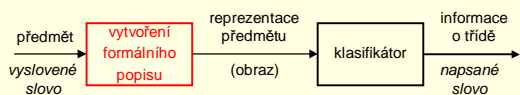


6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky

12

Reprezentace řeči pomocí příznaků



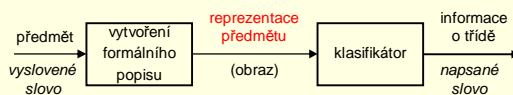
Vhodnými příznaky pro systémy rozpoznávání řeči jsou frekvenční charakteristiky řečového signálu.

6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky

13

Návrh klasifikátoru pro rozpoznávání řeči



Při návrhu klasifikátoru pro rozpoznávání řeči je třeba řešit celou řadu problémů.

6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky

14

Hlavní problémy při rozpoznávání řeči

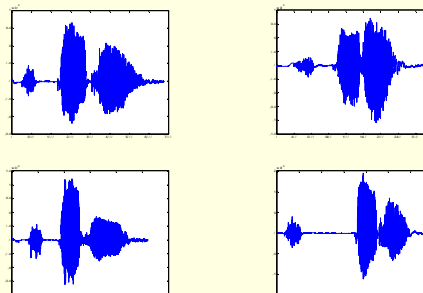
- n tutěž promluvu vysloví každý řečník jinak
- n dokonce stejný řečník vysloví tutěž promluvu pokaždě jinak

6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky

15

Příklad (tentyž řečník, nahráno v rozmezí 5 minut)



6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky

16

Hlavní problémy při rozpoznávání řeči

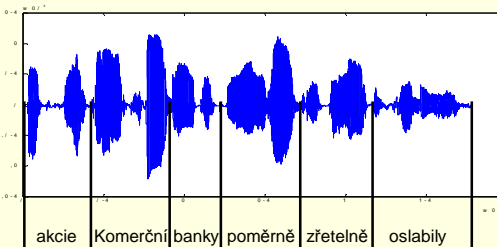
- n tutěž promluvu vysloví každý řečník jinak
- n dokonce stejný řečník vysloví tutěž promluvu pokaždě jinak
- n při rozpoznávání plynulé řeči nelze jednoduše automaticky stanovit začátek a konec jednotlivých slov

6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky

17

Příklad – zkuste najít hranice slov



6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky

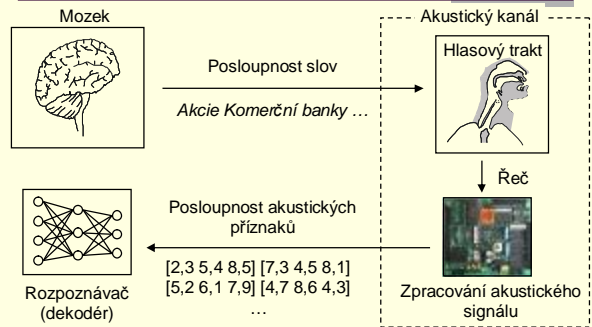
18

Hlavní problémy při rozpoznávání řeči

- n tutěž promluvu vysloví každý řečník jinak
- n dokonce stejný řečník vysloví tutěž promluvu pokaždě jinak
- n při rozpoznávání plynulé řeči nelze jednoduše automaticky stanovit začátek a konec jednotlivých slov

Téměř všechny výše zmíněné problémy se dají vyřešit nebo alespoň zmírnit statistickým přístupem

Model pro rozpoznávání (souvislé) řeči

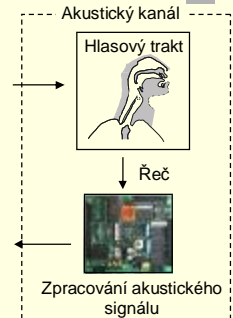


Rozpoznávání souvislé řeči

- n Klasifikátor pro rozpoznávání souvislé řeči (dekodér) je tvořen 2 částmi:
 - n akustickým modelem
 - n jazykovým modelem

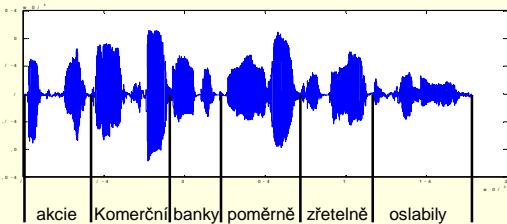
Akustický model

- n Modeluje akustický kanál, tj. zachycuje jak vlastnosti hlasového traktu, tak i charakteristiky bloku zpracování signálu (mikrofonu, zvukové karty...).



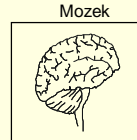
Akustický model

- n Jeho parametry se nastavují automaticky na základě trénovací množiny, která obsahuje nahrané věty spolu s jejich přepisem do textové podoby.



Jazykový model

- n Určuje pravděpodobnost, že si řečník bude přát vyslovit danou posloupnost slov
- n Jeho parametry se opět nastavují automaticky na základě trénovací množiny – v tomto případě stačí text, ve kterém se (zjednodušeně řečeno) zjišťují počty výskytů jednotlivých slovních n-tic



Jazykový model

- Při rozpoznávání pak stačí, když pro určení následujícího slova využijeme informaci o několika slovech předchozích.
- Příklad části předpovědi počasí:

... tlaková tendence: slabý *pokles*
vzestup

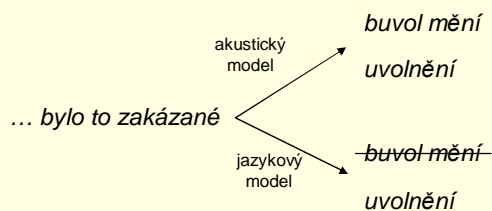
6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd,
katedra kybernetiky

25

Rozpoznávání souvislé řeči

- Využívá se kombinace výstupů akustického a jazykového modelu.
- Příklad části komentáře hokejového utkání:



6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd,
katedra kybernetiky

26

Rozdělení systémů rozpoznávání souvislé řeči

- podle způsobu řeči
 - rozpoznávání izolovaných slov
 - rozpoznávání souvislé řeči
- podle velikosti slovníku
 - systémy s malým slovníkem (hlasové povely)
 - systémy se středně velkým slovníkem (omezené úlohy – předpověď počasí)
 - systémy s velkým slovníkem (lze hovořit přirozeně na jakékoli téma)
- podle počtu uživatelů
 - systémy závislé na řečníkovi (1 uživatel systému)
 - systémy nezávislé na řečníkovi (více uživatelů)

6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd,
katedra kybernetiky

27

Možnosti využití systémů rozpoznávání řeči

- přepis řeči do textu
- titulkování televizních pořadů (skryté titulky)
- vyhledávání informací v rozsáhlých řečových archivech
- vyhledávání klíčových slov v proudu řeči
- hlasové dialogové systémy

6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd,
katedra kybernetiky

28

On-line titulkování sportovních přenosů



6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd,
katedra kybernetiky

29

On-line titulkování zasedání parlamentu



6.3.2009

Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd,
katedra kybernetiky

30

Neuronové sítě pro humanitní studia – rozpoznávání mluvené řeči

Hledání klíčových slov v záznamech schůzí Parlamentu ČR

6.3.2009 Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky 31

Neuronové sítě pro humanitní studia – rozpoznávání mluvené řeči

Dialogový systém „Přijímací řízení“

```

graph TD
    Start(( )) --> A[Pozdrav a informace]
    A --> B[Vlož své IČ]
    B --> C{IČ nalezeno?}
    C -- - --> D[Vybrat fakultu ze seznamu]
    C -- + --> B
    D --> E[Vybrat studijní obor ze seznamu]
    E --> F[Výsledky zkoušky]
    F --> G{Další info?}
    G -- + --> B
    G -- - --> H(( ))
  
```

6.3.2009 Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky 32

Neuronové sítě pro humanitní studia – rozpoznávání mluvené řeči

Příklady použití hlasových dialogových systémů

- n Přihlašování na zkoušky – alternativa k přihlašování na zkoušky přes internet, tel. 377 638 801)
- n Automatická spojovatelka, tel. 377 631 111
- n Aktuální zpravodajství – čtení aktuálních novinek (zpráv) ze serveru www.idnes.cz pomocí formátu RSS, tel. 377 638 805
- n Výsledky přijímacího řízení na ZČU přes telefon – v provozu již od roku 2000, každý rok (červen, červenec) obslouží 2000-3000 hovorů
- n Více informací na voice.zcu.cz

6.3.2009 Vlasta Radová, Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky 33