

Semestrální zkouška ZRE, řádný termín, 12.5.2008, skupina A

Login:

Podpis:

1. Blok normalizace textu v syntezátoru řeči převede řetězec "123,456" na

A: 3 jednotky: "123" ", " "456"

B: řetězec "sto dvacet tři celé čtyři sta padesát šest tisícin"

C: morfologickou informaci ke každé obsažené číslovce

D: sekvenci značek pro určení průběhu základního tónu.

2. Je v českém systému pro syntézu řeči z textu nutná desambiguace homografů (česky "zjednoznačňování stejně napsaných slov") ?

A: ne — dostaneme kvalitní převod na fonémy, prozódie může utrpět kvůli nepřesnému určení funkce slova ve větě.

B: ne — dostaneme kvalitní převod na fonémy, prozódie bude v pořádku (nepřesné určení funkce slova ve větě není možné).

C: ano, v češtině je přepis na fonémy určen významem.

D: ano, bez desambiguace homografů nelze použít metodu PSOLA.

3. Jaký je hlavní rozdíl mezi PCA (principal component analysis) a LDA (linear discriminant analysis) ?

A: LDA předpokládá přesné ohraničení vektorů z jednotlivých tříd, PCA ne.

B: LDA hledá směry podle maximální diskriminability tříd, PCA podle maximální variance dat.

C: LDA nepotřebuje mít definované třídy, PCA ano.

D: PCA poskytuje de-korelované příznaky, LDA nikoliv.

4. Blok RASTA filtrace je ve výpočtu parametrů v některých rozpoznávacích používán proto, že:

A: snižuje výpočetní náročnost rozpoznávání (omezení počtu koeficientů a možnost prodloužení analyzačního okna → méně rámců za stejnou jednotku času).

B: Zaručuje, že výsledné parametry budou mít gaussovské rozložení.

C: Omezí vliv událostí, které nemohly být vyprodukovány lidským řečovým ústrojím.

D: Odfiltruje z řečového signálu možné zbytky brumu (50 Hz rušení).

5. Polohu formantů F_1 a F_2 u jednotlivých samohlásek určuje:

A: frekvence kmitání hlasivek.

B: intenzita proudu vzduchu z plic.

C: národnost a psychický stav mluvčího.

D: poloha artikulačního ústrojí, především jazyka.

6. U systémů rozpoznávání s velkým slovníkem se slova nejčastěji:

A: skládají z menších jednotek: každý foném nebo foném závislý na kontextu je representován jednou Gaussovskou.

B: skládají z menších jednotek: každý foném nebo foném závislý na kontextu je representován jedním skrytým Markovovým modelem.

C: skládají z menších jednotek: každý foném nebo foném závislý na kontextu je representován rozpoznávací sítí udávající jeho výslovnostní varianty.

D: reprezentují nezávislými skrytými Markovovými modely, obvykle se 17ti stavy.

7. Jazykový model produkuje tzv. **apriorní** pravděpodobnost sekvence slov $\mathcal{P}(W_1^N)$. To znamená, že pro odhad této pravděpodobnosti:

- A: Je nutné znát počet vstupních akustických vektorů.
 - B: Je nutné znát všechny vstupní akustické vektory.
 - C: Je nutné znát všechny vstupní akustické vektory a navíc ještě identitu mluvčího.
 - D: Není nutné vědět o vstupních akustických vektorech vůbec nic.
-

8. Trénovací sekvence má 4 vektory: $\mathbf{o}(1) = [0.5; 1]^T$, $\mathbf{o}(2) = [1.2; 3]^T$, $\mathbf{o}(3) = [1; 2]^T$, $\mathbf{o}(4) = [-0.5; -3]^T$ a hodnoty state occupation function pro druhý stav HMM jsou: $L_2(1) = 0.5$, $L_2(2) = 0.4$, $L_2(3) = 0.2$, $L_2(4) = 0.1$. Určete nový odhad střední hodnoty druhého stavu $\boldsymbol{\mu}_2$:

- A: $\boldsymbol{\mu}_2 = [0.55; 0.75]^T$
 - B: $\boldsymbol{\mu}_2 = [0.73; 1.5]^T$
 - C: $\boldsymbol{\mu}_2 = [0.88; 1.8]^T$
 - D: $L_2(t)$ jsou špatně, protože jejich součet není 1.
-

9. Skrytý Markovův model má 4 stavy, z toho 2 vysílací. Jeho matice logaritmičeských přechodových pravděpodobností je:

$$\mathbf{A} = \begin{bmatrix} -\infty & 0 & -\infty & -\infty \\ -\infty & -0.11 & -2.30 & -\infty \\ -\infty & -\infty & -0.22 & -1.60 \\ -\infty & -\infty & -\infty & -\infty \end{bmatrix}$$

Model má na vstupu sekvenci 50ti vektorů. Hodnoty tokenů ve stavech 2 a 3 pro $t = 11$ jsou: $token_2(11) = -5$ a $token_3(11) = -6$. Jaká bude hodnota tokenu ve stavu 3 v čase $t = 12$, je-li hodnota log-vysílací pravděpodobnosti $\log b_3(\mathbf{o}(12)) = -0.54$?

- A: stav 3 nebude v tomto čase žádný token obsahovat.
 - B: $token_3(12) = -6.22$
 - C: $token_3(12) = -6.76$
 - D: $token_3(12) = -14.06$
-

10. Model M má tuto matici přechodových pravděpodobností:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

a zpracovává sekvenci sekvenci 4 vektorů. Výsledné vysílací pravděpodobnosti jsou:

	$\mathbf{o}(1)$	$\mathbf{o}(2)$	$\mathbf{o}(3)$	$\mathbf{o}(4)$
b_2	0.0349	0.0398	0.0013	0.0001
b_3	0.0010	0.0033	0.0340	0.0129

Určete pravděpodobnost vyslání těchto 4 vektorů modelem po stavové sekvenci: $X = [1 \ 2 \ 2 \ 2 \ 3 \ 4]$

- A: $P(\mathbf{O}, X|M) = 6.71 \times 10^{-10}$
- B: $P(\mathbf{O}, X|M) = 7.85 \times 10^{-10}$
- C: $P(\mathbf{O}, X|M) = 1.72 \times 10^{-11}$
- D: $P(\mathbf{O}, X|M) = 4.40 \times 10^{-11}$

-
11. Skrytý Markovův model zpracovává 69-rozměrné vektory¹. Vysílací funkce hustoty rozdělení pravděpodobnosti jednoho ze stavů je určena jednou 69-rozměrnou Gaussovskou s parametry $\boldsymbol{\mu} = [1, 1, \dots, 1]^T$ a $\boldsymbol{\sigma} = [4, 4, \dots, 4]^T$ (všechny střední hodnoty a všechny směrodatné odchylky jsou tedy stejné). Jaká je hodnota této funkce hustoty pro vstupní vektor $\mathbf{o} = [1, 1, \dots, 1]^T$.

- A: $b(\mathbf{o}) = 0.0997$
 B: $b(\mathbf{o}) = 6.11 \times 10^{-30}$
 C: $b(\mathbf{o}) = 1.52 \times 10^{-49}$
 D: $b(\mathbf{o}) = 8.33 \times 10^{-70}$
-

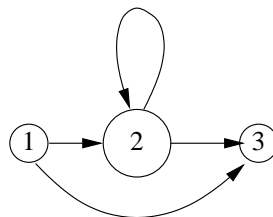
12. U HMM s plnými kovariančními maticemi je vysílací pravděpodobnost:

$$b_j[\mathbf{o}(t)] = \mathcal{N}(\mathbf{o}(t); \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) = \frac{1}{\sqrt{(2\pi)^P |\boldsymbol{\Sigma}_j|}} e^{-\frac{1}{2}(\mathbf{o}(t) - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{o}(t) - \boldsymbol{\mu}_j)}$$

Uveďte, kdy je nutné vyhodnotit determinant kovarianční matice a provést její inverzi.

- A: obojí pouze po odhadu nové kovarianční matice (rovnice (10) ve slajdech o HMM).
 B: inverze pouze jednou po odhadu kovarianční matice, determinant pro každý vektor $\mathbf{o}(t)$.
 C: determinant pouze jednou po odhadu kovarianční matice, inverze pro každý vektor $\mathbf{o}(t)$.
 D: inverze i determinant pro každý vektor $\mathbf{o}(t)$.
-

13. V rozpoznávání s velkým slovníkem se za každé slovo přidává takzvaný “tee-model” krátkého ticha (sp – short pause), který vypadá takto:



Proč má tuto strukturu ?

- A: Dovede zpracovat více vektorů ticha než běžné modely bez skoků.
 B: Skok je nutný pro zahrnutí vlivu hodinové frekvence CPU.
 C: Ve spojitě řeči může za slovem následovat ticho, ale také nemusí.
 D: Ve spojitě řeči jsou slova vždy oddělena tichem.
-

14. Je dána “mřížka” lokálních vzdáleností $d(n, m)$ (reference svisle, test vodorovně):

$$\begin{bmatrix} 1 & 5 & 3 \\ 5 & 2 & 1 \\ 4 & 3 & 6 \\ 3 & 2 & 5 \end{bmatrix}$$

Jaká je DTW vzdálenost ?

- A: 1.7
 B: 2.0
 C: 2.3
 D: 2.9

¹Toto číslo jsem si tak úplně nevycucal z prstu, systém Speech@FIT pro evaluace rozpoznávání holandštiny 2008 má skutečně tento počet parametrů!

-
15. Určete Euklidovu vzdálenost dvou tříprvkových vektorů: $\mathbf{x}_1 = [-1.5, 2.1, 7.2]^T$,
 $\mathbf{x}_2 = [-2.1, 1.5, 6.3]^T$

A: $d(\mathbf{x}_1, \mathbf{x}_2) = 1.24$

B: $d(\mathbf{x}_1, \mathbf{x}_2) = 1.53$

C: $d(\mathbf{x}_1, \mathbf{x}_2) = [-0.6, -0.6, -0.9]^T$

D: $d(\mathbf{x}_1, \mathbf{x}_2) = [0.36, 0.36, 0.81]^T$

16. U srovnání metodou DTW má reference délku R a test délku T . Co platí o indexovacích sekvencích $r(k)$ a $t(k)$ pro délku cesty K ?

A: $r(K) = K, \quad t(K) = K$

B: $r(K) = \max(K - 1, R, T), \quad t(K) = \max(K + 1, R, T)$

C: $r(K) = R, \quad t(K) = T$

D: $r(k) + t(k) = T + R$ pro libovolné $k \in [1, T + R]$

17. V CELP kódérech je dlouhodobý prediktor nahrazen tzv. adaptivní kódovou knihou, která obsahuje "historii buzení". Proč tedy GSM-EFR (ETSI 06.60) obsahuje detekci lagu, když by se dalo pouze vyhledávat v adaptivní kódové knize ?

A: tato detekce lagu slouží pouze pro detekci znělosti, pak je zapomenuta.

B: tato detekce vybere kandidáty lagu, okolo nich se pak vyhledává v adaptivní kódové knize (urychlení).

C: z takto detekovaného lagu jsou určeny LPC koeficienty.

D: z takto detekovaného lagu se určuje frekvenční charakteristika perceptuálního filtru.

18. Proč je v klasické telefonní PCM použito logaritmické kvantování (resp. aproximace pomocí A-law nebo μ -law) ?

A: v době vzniku normy G711 nebylo lineární kvantování známo.

B: 8bit logaritmické kvantování je pro neřečové signály (např. bílý šum) kvalitnější než 16bit lineární.

C: koriguje se pomocí něj charakteristika ucha, která je exponenciální.

D: pro řečové signály se 8bit logaritmickým kvantováním dosáhne podobné kvality jako se 13bit lineárním.

19. Jak by měl ve vokodéru vypadat optimální tvar jednoho budícího pulsu pro znělé signály ?

A: Mohl by být získán ze skutečné řeči inverzním filtrem $A(z)$.

B: Impuls s hodnotou 1, ostatní vzorky 0.

C: Část cosinusovky na hodnotě $3 \times F_0$ (frekvence základního tónu).

D: několik vzorků bílého šumu.

20. Proč je chybový signál lineární predikce (tzv. LPC residual) vhodný pro určování základního tónu ?

A: Frekvence základního tónu jsou v něm omezeny do intervalu [20 Hz, 160 Hz].

B: Tento signál neobsahuje spektrální složky nad 2500 Hz (pro telefonní signál se vzorkovací frekvencí 8000 Hz).

C: LPC residual nemá stejnosměrnou složku.

D: LPC residual neobsahuje informaci o formantech.