

Semestrální zkouška ZRE, řádný termín, 13.5.2022, skupina Vishnu

Login: Příjmení a jméno: Podpis:
(prosím čitelně!)

1. Nakreslete blokové schéma výpočtu Mel-cepstra.

2. Napište definici poměru signálu k šumu v dB a vysvětlete, co je co.

3. Popište princip diferenční pulsně kódové modulace (DPCM) - schéma + kratičký text. Pro zjednodušení nemusíte uvažovat kvantování chybového signálu.

4. Pro kódování CELP je klíčový princip analýzy syntézou (analysis by synthesis). Vysvětlete ho, nejlépe schéma + kratičký text.

5. Proč je reziduální signál krátkodobé lineární predikce (např. s $P = 10$) stále nevhodný ke kódování a co se s tím v současných kodérech dělá ?

6. Popište vztah mezi dlouhodobým prediktorem a adaptivní kódovou knihou v kóděrech typu CELP.

7. Referenční slovo je “páv” (0.5 s) a testovací slovo je “paf” (0.25 s), obě vyslovil stejný mluvčí ve stejných podmínkách. Nakrelete, jak bude vypadat optimální srovnávací cesta DTW, referenci kreslete svisle, test vodorovně. Vektory parametrů jsou extrahovány s krokem 10 ms.

8. Napište, jak vyhodnotit věrohodnost (likelihood) sekvence T feature vektorů $\mathbf{O} = [\mathbf{o}_1 \dots \mathbf{o}_T]$ pomocí modelu směsi Gaussovských rozložení (GMM) s následujícími parametry: J gaussovských komponentů, každý má střední hodnotu $\boldsymbol{\mu}_j$, kovarianční matici $\boldsymbol{\Sigma}_j$ a váhu w_j . Stačí základní vzorec (bez expanze, jak funguje vícedimenzionální Gaussovka).

9. Je dán HMM s $N = 4$ stavy, z toho 2 jsou vysílací. Nenulové přechodové pravděpodobnosti jsou pouze $a_{12} = 1$, $a_{22} = 0.6$, $a_{23} = 0.4$, $a_{33} = 0.7$, $a_{34} = 0.3$, ostatní jsou nulové. Funkce hustoty rozdělení pravděpodobnosti v jednotlivých stavech jsou dány jedním Gaussovým rozložením s následujícími vektory středních hodnot $\boldsymbol{\mu}_2 = [1 \ 2]^\top$, $\boldsymbol{\mu}_3 = [-1 \ -2]^\top$ a oba mají stejný vektor směrodatných odchylek: $\boldsymbol{\sigma}_2 = \boldsymbol{\sigma}_3 = [2 \ 2]^\top$.

Vstupní sekvence má 5 vektorů: $\mathbf{O} = \left[\begin{array}{c|c|c|c|c} 0.5 & 1.2 & 1 & -0.5 & -1 \\ \hline 1 & 3 & 2 & -3 & -5 \end{array} \right]$. Napište, jaká bude pro tento vstup stavová sekvence s nejvyšší věrohodností a velmi krátce zdůvodněte. Help: tato sekvence bude mít délku 7.

10. Je definován levo-pravý HMM se čtyřmi stavy, z toho 2 vysílací, přechodové log. pravděpodobnosti jsou: $\log a_{12} = 0$, $\log a_{22} = -0.51$, $\log a_{23} = -0.92$, $\log a_{33} = -0.36$, $\log a_{34} = -1.2$. Tabulka logaritmů hodnot funkcí hustoty vysílacích likelihoodů je:

t	...	46	47	48	...
$\log b_2(\mathbf{o}_t)$	...	-1	-2	-3	...
$\log b_3(\mathbf{o}_t)$	...	-4	-5	-6	...

Provádíme Viterbiho algoritmus pomocí “token passing”. Hodnota tokenu ve stavu 2 v čase 46 je $\Psi_2(46) = -21$. Určete hodnotu tokenu ve stavu 2 v čase 48.

$\Psi_2(48) = \dots\dots\dots$

11. Jak aproximujeme pravděpodobnost sekvence slov $P(w_1, w_2 \dots w_N)$ v případě trigramového jazykového modelu ?

12. Formulujte problém rozpoznávání řeči s velkým slovníkem (LVCSR) pravděpodobnostně pomocí Bayesova vzorce. Krátce vysvětlete, co jsou jednotlivé pravděpodobnosti, resp. likelihoody.

13. Při rozpoznávání řeči pomocí váhovaných konečných stavových akceptorů, resp. převodníků (wFSA a wFST) je výsledná rozpoznávací síť dána jako $HCLG = H \circ C \circ L \circ G$. Vyberte si jeden z komponentů a popište, co jsou vstupní symboly, co jsou pravděpodobnosti, a pokud je to wFST, co jsou výstupní symboly.

14. Co je to posteriogram ? Nakrelete příklad pro slovo “sen”.

15. Co je na výstupu Connectionist Temporal Classification (CTC) ?

16. Nakreslete obecné schéma sequence-to-sequence modelování a uveďte, jak se dá aplikovat na LVCSR.

17. Jak se vyhodnocuje ověřování mluvčího (speaker verification) ? Help: stejně jako jakákoliv detekční úloha.

18. Popište architekturu pro extrakci speaker-embeddingů typu x-vektor pro ověřování mluvčího. Pokud ji neznáte, vymyslete ji. Help: embedding musí mít pro promluvu jakékoliv délky stejný počet koeficientů.

19. Co je v syntéze z textu (TTS) normalizace textu ? Uveďte 3 příklady.

20. Srovnajte klasický přístup k TTS s end-to-end přístupem. Můžete např. nakreslit schémata a doplnit kratičkým textem.