Introduction to Digital Speech Processing, Final Exam

Jun. 24, 2015, 10:10-12:10

- OPEN Lecture Power Point (Printed Version) and Personal Notes
- You have to use CHINESE sentences to answer all the questions, but you can use english terminologies
- Total points: 170

---

1. (20) (1) Describe how the MAP principle can be used for speaker adaptation. (2) Describe how Maximum Likelihood Linear Regression (MLLR) can be used for speaker adaptation.
2. (20) (1) Describe what Cepstral Mean Subtraction(CMS) and Cepstral Mean and Variance Normalization(CMVN) are in robustic speech recognition against environmental mismatch. (2) Describe what Histogram EQualization(HEQ) is in robustic speech recognition.
3. (20) (1) Describe what Markov Decision Process(MDP) is including its fundamental mathmatical framework. (2) Describe how MDP can be used in user-content interaction and spoken dialogue systems.
4. (10) Please explain what is the "lattice" of an utterance, and what is the "expected term frequency" for a term in the lattice.
5. (10) Explain what is Psueudo-Relevance Feedback (PRF) and its main procedure in spoken document retrieval.
6. (10) Explain what is Maximum Margin Relevance (MMR) in spoken document summarization and describe how it works in detail.
7. (20) Explain briefly what is Latent Semantic Analysis (LSA) and how Singular Value Decomposition (SVD) works in LSA.
8. (20) What is WFST? How is it useful in speech recognition?
9. (10) What is the Random Forest? How to use Random Forest for Tone Recognition?
10. (20) What is the EM algorithm? Explain how it can be used to solve the Basic Problem 3 for a discrete HMM.
11. (10) Explain what is the Conditional Random Field (CRF) and how it can be used for slot filling in spoken dialogues.

1.(ch13 p.4~p.10) +(中文講義)
(1)
MAP(Maximum A Posterior)收集到speaker的data之後，根據現有的model set，調整出現過的model的mean值，使該model set對於給定data的機率最大化。
(2)
MAP只能調整有出現過的model，需要大量data才能train得好，MLLR將model分成好幾個class，整個class一起調整，在data量少的時候效果較好。
以上兩種方法都是用EM algorithm去調整參數。

2.(ch15 p.9~p.11)
(1)
CMS是為了消除convolutional noise，從time domain轉換到cepstral domain，使noise可以用減法消除，用EM algorithm求要調整的量。CMVN是在CMS之後多做一次normalization，使形狀跟本來的更像。
(2)
HEQ是將原distribution的cumulative probability直接對應到另一個distribution。

3.(ch11,ch12)
(1)
Markov Decision Process (MAP) is mathematical framework for decision making
it is defined by a 5-tuple (S,A,T,R,π) where
S is the set of states, current system status
A is the set of actions the system can take at each state
T transition probabilities between states where a certain action is taken
R reward received when taking an action
π is the policy or choice of action given the state

the objective of MDP is to find a policy that maximizes the expected total reward
(2)
in user content interaction:
當一個查詢被輸入,系統從某個狀態(state)開始運作

states:檢索回傳的結果以某個連續變數(如MAP)估計的值加上目前對話的回合數

action:系統在每個狀態都有一些動作可以選擇:詢問更多資訊,回傳一個關鍵字或,
回傳一串關鍵字或文檔請使用者選一個,或顯示結果

user response corresponds to a certain negative reward(extra work for user)

when the system decides to show to the user the retrieved results, it earns some positive reward (e.g. MAP improvement)
Learn a policy maximizing rewards from historical user interactions ($\pi: S_i \rightarrow A_j$)
in spoken dialogue systems:
參考ch17投影片 飛機語音訂票系統 or 自由發揮厂

4.(ch10)

lattice:

辨識語音時將每個發音的前N個最佳結果串在一起形成的網格

expected term frequency:

公式：$E(t,x) = \Sigma N(t,u)*P(u|x)$ -> summation over 每一條包含x的sequence

u: lattice 中經過某個utterence的一條路徑(sequence)

P(u|x): 給定 utterance x 得到一條 sequence u的事後機率

N(t,u): term t在 sequence u 裡面出現的次數

L(x): lattice 中所有包含 utterance x 的路徑

(2016)8、

5.(ch10)

pseudo relevance feedback:

一種自動產生訓練資料的方法

流程：

先進行一次first pass retrieval 產生結果

假設前n個搜尋結果與查詢目標有相關

假設後m個搜尋結果與目標無關

重新計算每個結果與前兩個步驟找出的pseudo relevant/irrelevant utterance的acoustic similarity進行re-ranking

把re-rank出來的結果回傳給user

acoustic similarity是用dynamic time warping算出來的

---

6 (ch11):

MMR是產生summary的方法。

MMR(Xi) = Rel(Xi, D) – lambda * Red(Xi, S)

Rel(Xi, D) = sim(Xi, D)

Red(Xi, S) = sim(Xi, S)

X為句子，D為文件，S為已生成的summary, Sim為相似度

每次從文件中挑MMR值最大的一句，加入summary中，為了避免重複概念太多句(同樣的概念會出現很多次，然後都跟文件很像)，所以逐步加大lambda, 避免重複產生和summary相似的句子。

7(ch 14).

找出潛藏在文章中的語意。可以用來找出文章的主題。例如兩篇文章都有固定的哪些詞，就可能是相同的主題。

一種方法是算出每篇文章word的distribution, 若是兩篇文章的distribution相似，則很有可能是相同的主題。

但是在文章數很大的時候，矩陣的維度會變得很大，SVD是一種降維的方法，可以保有原本的dsitribution的狀況下降低維度。

$$W_{M\times N} \approx \hat{W}_{M\times N} = U_{M\times R} S_{R\times R} V^T_{R\times N}$$

s_i: singular values, $s_1 \geq s_2 \ldots \geq s_R$

把矩陣拆解成3個相乘, 中間的那個稱為singular value, 因為會愈來愈小, 可以只取前幾個, 只算出W的一部份, 就可以達到降維的效果。就可以很快看出這個文件的相似度。

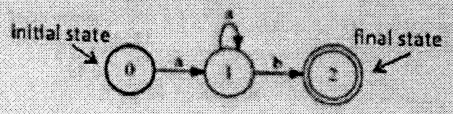
==========

8. (ch 9)p.58 (2016) 9、

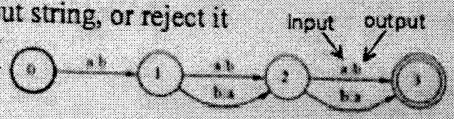
Weighted Finite State Transducer

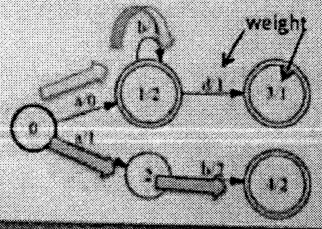
Weighted Finite State Transducer(WFST)

- **Finite State Machine**
  - A mathematical model with theories and algorithms used to design computer programs and digital logic circuits, which is also called "Finite Automaton".
  - The common automata are used as acceptors, which can recognize its legal input strings.
- **Acceptor**
  - Accept any legal string, or reject it
  - EX: {ab, aab, aaab, . . .} = aa*b



- **Transducer**
  - A finite state transducer (FST) is a extension to FSA
  - Transduce any legal input string to another output string, or reject it
  - EX: {aaa, aab, aba, abb} -> {bbb, bba, bab, baa}
- **Weighted Finite State Machine**
  - FSM with weighted transition
  - An example of WFSA
  - Two paths for "ab"
    - Through states (0, 1, 1); cost is (0+1+2) = 3
    - Through states (0, 2, 4); cost is (1+2+2) = 5





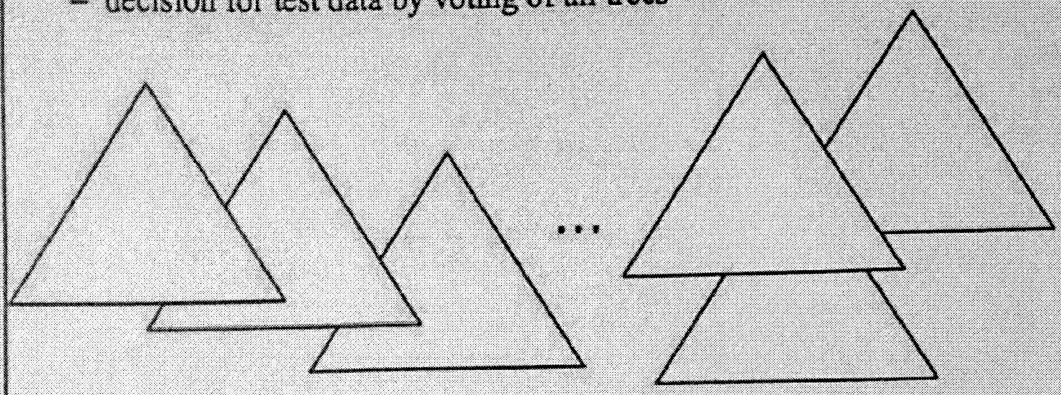
√ **how is it useful in speech recognition?**

WFST 的好處為可以經由最佳化使搜尋空間縮小,減少辨識所需的時間,另 外,WFST 皆以相同表達方式表示各個層級,各個模型皆以狀態和轉移表示,方 便系統以標準化的方式處理。

9.(ch9)p.70

Random Forest for Tone Recognition for Mandarin

- **Random Forest**
  - a large number of decision trees
  - each trained with a randomly selected subset of training data and/or a randomly selected subset of features
  - decision for test data by voting of all trees



==========

10.(ch16)

1)

EM algorithm 是在機率模型當中，找到參數使得maximum a posteriori 能夠最大化，而在機率模型當中包含了無法觀測的latent variable。每個iteration的過程包括了E-step和M-step，E-step: 根據現有的data和參數得到一個latent data 的distribution後，來表示MAP(objective function)的期望值。M-step: 更新參數使得MAP(objective funciton)的值變大。

2)

Example: Use of EM Algorithm in Solving Problem 3 of HMM

- **Observed data : *observations* O, latent data : *state sequence* q**
- **The probability of the complete data is**
  $P(O,q|\lambda)= P(O|q,\lambda)P(q|\lambda)$
- **E-Step :**
  $Q(\lambda, \lambda^{[k]})=E[\log P(O,q|\lambda)|O, \lambda^{[k]}]= \Sigma_q P(q|O,\lambda^{[k]})\log[P(O,q|\lambda)]$
  - $\lambda^{[k]}$: k-th estimate of λ (known), λ: unknown parameter to be estimated
- **M-Step :**
  - Find $\lambda^{[k+1]}$ such that $\lambda^{[k+1]} =\arg\max_{\lambda} Q(\lambda, \lambda^{[k]})$
- **Given the Various Constraints** (e.g. $\sum_i \pi_i = 1, \sum_j a_{ij} = 1$, *etc.*), **It can be shown**
  - the above maximization leads to the formulas obtained previously
  - $P(O|\lambda^{[k+1]}) \geq P(O|\lambda^{[k]})$

11.(ch17)

1)

CRF算是一種Machine learning的方法。Input 一段observation sequence x，會希望ouput 出一段label sequence y。而這些output 出來的label sequence，比起observation sequence, 關係是比較明確的。(ex: input: "Amy ate lunch at KFC"; output label: "Noun Verb Noun Preposition Noun")

2)

以搜尋電影為例，可以label input sentence 演員 劇情等，因此當使用者說完一句話後就能找到關鍵的字詞。