Corrigé Exam M2 APAS 1

11- 2007

1.

1.1 $P(F,A|E) = \prod_{j=1}^{m} p(f_j, a_j | f_1^{j-1}, a_1^{j-1}, E)$

$= \prod_{j=1}^{m} p(a_j | a_1^{j-1}, f_1^{j-1}, E) \cdot p(f_j | f_1^{j-1}, a_1^{j}, E)$

1.2 Quand on génère 1 phrase F et son alignement A

- on choisit à quel ~~mot de E on connecte le 1er mot de F~~ position dans E connecter la 1ère position dans la phrase Française sachant E et
- on choisit lequel 1er mot de F connaissant E et la position a_1 etc

2

2-1

$P(a_j | a_1^{j-1}, f_1^{j-1}, E) = \frac{1}{n+1}$: o proba. uniforme o ne dépend que de n o ind. des a_j préc. E, F

$P(f_j | a_1^{j}, f_1^{j-1}, E) = P(f_j | e_{a_j})$: f_j ne dépend que de la position a_j et du mot e_{a_j} /pas du passé-

2.2

$$P(F,A|E) = \frac{1}{(n+1)^m} \prod_{j=1}^{m} P(f_j|e_{a_j})$$

2.3

$$P(F|E) = \sum_A P(F,A|E)$$
$$= \frac{1}{(n+1)^m} \sum_{a_1=0}^{n} \cdots \sum_{a_m=0}^{n} \prod_{j=1}^{m} P(f_j|e_{a_j})$$

3

3.1 Melange de densités / HMM

3.2 Non → equation couplées

3.3 $Q(\theta,\hat{\theta})$ est l'espérance de la log vraisemblance $\log P_\theta(F,A|E)$ pour la distribution de A

3.4

$$Q(\theta,\hat{\theta}) = \sum_A \left(\log P_\theta(F,A|E)\right) P_{\hat{\theta}}(A|F,E)$$
$$= \sum_A \left(\log \frac{1}{(n+1)^m} + \log \prod_{j=1}^{m} P_\theta(f_j|e_{a_j})\right) P_{\hat{\theta}}(A|F,E)$$

$$\boxed{Q(\theta,\hat{\theta}) = \sum_A \left(\log \prod_{j=1}^{m} P_\theta(f_j|e_{a_j})\right) P_{\hat{\theta}}(A|F,E) + \sum_A \log \frac{1}{(n+1)^m} P_{\hat{\theta}}(A|F,E)}$$

3.5

$$\mathcal{L} = Q(\theta, \hat{\theta}) - \sum_e \lambda_e \sum_f (P_\theta(f|e) - 1)$$

3.6

$$\frac{\partial \mathcal{L}}{\partial P_\theta(f|e)} = \frac{\partial Q}{\partial P(f|e)} - \lambda_e$$

$$= \frac{\partial}{\partial_\theta P(f|e)} \left(\sum_{j=1}^{m} \sum_{k=1}^{n} P_{\hat{\theta}}(k|j, F, E) \log P_\theta(f_j|e_k) + K \right) - \lambda_e$$

$$= \sum_{j=1}^{m} \sum_{k=1}^{n} P_{\hat{\theta}}(k|j, F, E)\, \delta(f, f_j)\delta(e, e_k) \cdot \frac{1}{P_\theta(f_j|e_k)} - \lambda_e$$

$$\frac{\partial \mathcal{L}}{\partial P_\theta(f|e)} = 0 \iff P_\theta(f|e) = \lambda_e^{-1} \sum_{j=1}^{m} \sum_{k=1}^{n} P_{\hat{\theta}}(k|j, F, E)\, \delta(f, f_j)\delta(e, e_k)$$

3.7

$$P_\theta(f|e) = \lambda_e^{-1} \sum_{j=1}^{m} \sum_{k=1}^{n} \frac{P_\theta(f_j|e_k)}{\sum_i P_\theta(f_j|e_i)} \times \delta(f,f_j)\,\delta(e,e_k)$$

$$(1) \quad \boxed{P_\theta(f|e) = \lambda_e^{-1} \frac{P_\theta(f|e)}{\sum_i P_\theta(f|e_i)} \sum_{j=1}^{m} \delta(f,f_j) \sum_{k=1}^{n} \delta(e,e_k)}$$

3.8 <u>Algorithme</u>

- initialiser les $P_\theta(f|e)$ $\forall f\ \forall e$
- E $\quad P_{\hat\theta}(k|j,F,E) = \frac{P_\theta(f_j|e_k)}{\sum_i P_\theta(f_j|e_i)}$ revoir étape
- M formule 1.

5 -   5

5 - 1 . CRF – méthode discriminative vs générative.
→ possibilité de prendre en compte des features arbitraires sur les entrées et les relations entrées sorties.
→ prise en compte de dépendances long terme.
° proba MV vs proba MAP
° régularisation

5 . 2 . features arbitraires
dép. long terme
dép. entrées / sorties

5 . 3 ° CRF prend en compte les dépendances entrées sortie
° régularisation pour éviter overfitting

5 . 4 h_k : features
λ_k : poids
Z : normalisation

5 . 5 données : phrases alignées
critère : MAP régularisé
algo : BFGS . gradient