Interactive Spoken Content Retrieval by Deep Reinforcement Learning

資工三 B05902118 陳盈如

Introduction

Interactive Information Retrieval (IIR)將user-system interaction和retrieval process兩者做結合，主要目的是為了讓使用者可以在互動之中找到所需的資訊，而user-system interaction如此重要的原因正是因為speech recognition至今仍會產生無可避免的錯誤，進而導致無法控制的結果。在此篇論文中提出Deep-Q-Network (DQN)來解決問題且有顯著的效果，另外，即使是使用raw feature當作DQN的input，所產生出來的結果仍優於原本的方法。

Proposed Approach

1. Retrieval Module:

- Language Modeling Retrieval Module:

$$S(q,d) = -[KL(\theta_q \| \theta_d) - \beta KL(\theta_N \| \theta_d)]$$

$S(q,d)$代表query q和document d之間的相對分數，$KL(\ )$用來計算兩個model之間的divergence

θ_q: query's language model、θ_d: documents's language model、θ_N: a set of terms which is not related.

- Query-Regularized Mixture Model for Query Expansion:

Query model有可能在query expansion process的時候受到一些不相關資訊的影響，故利用關鍵字去regularize query model。

2. System的五種actions:

照順序回應現有的document、詢問user關鍵字、要求user提供更多關鍵字、給幾個topic讓user選擇、給出所有的retrieval results並結束互動

3. **Feature Extraction:**

將feature分成兩種，Human Knowledge Feature和Raw Relevance Scores

4. Dialogue Manager:

- *Previous Approach: Estimating the hand-crafted state:*

此方法會分成兩個步驟，一是state estimation，二是action decision。在步驟一時，輸入上一階段得到的feature set並預測出average precision (AP)；步驟二則是在計算$Q(s,a)$，找出能使$Q(s,a)$最大化的action a。

缺點：

(1) AP並不能作為state的代表，即使AP相同，optimal action也有可能不同

(2) 分開train state estimation和action decision，故無法修正error margin

- *Proposed Approach: Deep Reinforcement Learning:*

  $\hat{y}_i = r + \gamma \cdot \max_{\alpha' \in A} Q(s', \alpha'; \theta^-)$, θ^-代表固定的目標network

  $L_i(\theta^i) = \mathbb{E}_{s,\alpha,r,s' \sim U(D)}[(\hat{y}_i - Q(s, \alpha; \theta^i))^2]$, $D = \{e_1, e_2, \ldots, e_L\}$, $e_t = (s_t, \alpha_t, r_t, s_{t+1})$

  DQN成功的關鍵是experience replay以及固定$Q(s', \alpha'; \theta^-)$，使$Q(s, \alpha; \theta^i)$穩定更新，透過hidden layers從feature產生最適合的action，也解決estimating hand-crafted state兩步驟分開train的問題。

Experiments

- **Experiment Setting:**

  利用台北2001~2003年所有中文的新聞的corpus作為target document，並且使用tri-gram language model。模擬使用者有以下四種回饋：從清單裡面選出最相關的document、關鍵字是否超過50%、輸入新的關鍵字、隨機回覆一個相關的topic。

- **Result and Discussion:**

Approaches	one-best		lattices	
	MAP	Return	MAP	Return
(a-1) First-pass	0.4521	-	0.4577	-
(a-2) Random Actions	0.4553	-61.7	0.4117	-111.21
(a-3) Hand-crafted States	0.5398	67.07	0.5626	84.54
(b-1) Raw Feature	0.5619	89.27	0.5847	105.03
(b-2) Selected Feature+(b-1)	0.5691	95.72	0.5907	110.90
(c) Upper Bound (Oracle)	0.6554	164.94	0.6639	168.75

(1) 隨機選取action的結果並不會優於hand-crafted states，因為即使每次迭代後能得到更多的資訊，但是更多次的request會使系統的負擔加重。

(2) 輸入raw feature當DQN的input得到的結果比原先方法使用raw relevance scores和human knowledge features所得到的結果還要好。

(3) 要從raw relevance scores得到action是困難且複雜的，故無法不使用hidden layers的方式。

Conclusion

這個end-to-end的學習方法可以使整體user-system interaction最佳化，相較之前的hand-crafted states也有顯著進步，甚至只有raw relevance的情況下效果都優於前者。

Reference

"Interactive Spoken Content Retrieval by Deep Reinforcement Learning", Interspeech, San Francisco, USA, Sept 2016.