HW3

r12922016 葉丞勛

Report – Canonical Knowledge

1. Logic Warm-up

   (1) 將每個 sentence 用 &, |, ~, % 等方式表示，最後用 conjoin 將他們變全部 and 起來。

   (2) findModelUnderstandingCheck()：利用一個 class，將他的 representation 設成回傳自己的一個型態為字串的屬性，如此一來，就能成功回傳 {a：True}。

   (3) plTrueInverse()：呼叫 pl_true()，當 assignments 帶入 ~inverse_statement 為 True 時，才回傳 True。

```
### Question q1: 10/10 ###


Finished at 22:26:09

Provisional grades
==================
Question q1: 10/10
------------------
Total: 10/10
```

2. Logic Workout

   (1) atLeastOne()：透過 disjoin 將所有 Expr 都 or 起來，當其中有至少一個為 True 的時候，就會回傳 True。

   (2) atMostOne()：由於最多只會有一個為 True，因此任兩個做 and 必定會是 False，因此我們可以將所有 Expr 兩兩一組，再將每組取 not，最後將他們透過 conjoin 將他們全部 and 起來即可。

   (3) exactlyOne()：將 atLeastOne 以及 atMostOne 做 and 即可。

```
### Question q2: 10/10 ###


Finished at 22:38:48

Provisional grades
==================
Question q2: 10/10
------------------
Total: 10/10
```

3. Pacphysics and Satisfiability

   (1) pacmanSuccessorAxiomSingle()：假設 pacman 在時間為 t 時處於 (x, y)，那麼 pacman 在 t-1 時，可以透過上下左右四種方式抵達(x, y)，我們可以透過 disjoin 將這四種方式 or 起來。

   (2) pacphysicsAxioms()：依照 readme 的做法，把所有條件列好後，透過 conjoin and起來。

   (3) checkLocationSatisfiability()：照 readme 的做法，把所有條件先加入 KB，接著將所有在 KB 的 Expr and 起來，接著令 condition 為 pacman 在 t=1 時在 (x1, y1)，最後透過 findModel(kb & condition)、findModel(kb & ~condition) 來取得兩個目標 model。

```
### Question q3: 10/10 ###


Finished at 23:05:15

Provisional grades
==================
Question q3: 10/10
------------------
Total: 10/10
```

4. Path Planning with Logic

   ：positionLogicPlan()：照著 readme 的做法，把全部條件加入 KB 後做 conjoin，接著設定 goalAssertion 為 pacman 在 time=t 時在 (xg, yg)，最後透過 findModel(kb & goalAssertion) 以及 extractActionSequence 取得一連串從 start 到 goal 的 actions。

```
### Question q4: 10/10 ###


Finished at 23:08:52

Provisional grades
==================
Question q4: 10/10
------------------
Total: 10/10
```

5. Eating All the Food

   ：foodLogicPlan()：整體做法與 Q4 相同，差別在於要加一個 Expr 進入 KB 裡面進行判斷：當 time=t 時, pacman 不在(x, y) 並且 food 在(x, y)時, 就代表在 time=t+1 時，food 會繼續在 (x, y) 上。 接著透過 ~disjoin 設定 goalAssertion 為當前所有的 food 都被吃光，最後透過 findModel(kb & goalAssertion) 以及 extractActionSequence 取得一連串會吃光所有 food 的 actions。

```
### Question q5: 10/10 ###


Finished at 23:15:37

Provisional grades
==================
Question q5: 10/10
------------------
Total: 10/10
```

Questions According to Lecture

Q1：

我認為 LLMs 之所以無法達到通用性(generality)的主要原因在於它只能夠被用來處理跟 Text 有關的任務，例如幫文章做重點摘要、算數學問題、回答問題等等，無法進行跟 Vision 相關的任務。因此， 若是人類在未來想要成功做出一個通用的 AI，那勢必得往多模態(multi-modal)的方向前進，讓 AI 能夠接受各種不同 type 的 input，以完成更多不同領域的任務。然而，現今的 VLMs 是以 object-centric 的方式去學習，而這種方法往往只能學習到圖片表面的意思而已，無法學習到這張圖背後抽象的意涵。並且，現今的 VLMs 對於取得 visual commonsense knowledge 的能力欠佳，實驗結果顯示：只有約莫 50% 的準確率可以成功取得圖片裡物體的顏色、大小、位置這些基本的資訊，若是想取得圖片裡物體跟物體間更複雜的資訊則更加困難。綜合上述，我認為要真的成功訓練出一個通用的 AGI，人類還有一段漫長的路要走。

Q2：

Positive samples：將一開始取得的 event-centric structured data 丟給 GPT，接著 GPT 會生出一段代表這個 event-centric structured data 的文字(t1)，而這段文字就是 positive samples。

Negative samples：在這篇 paper 的做法中，我們在意的是 hard negatives，也就是與 positive samples 只有細微差距的 samples。欲取得這類 sample，我們可以透過 (1) 只改變 structured data 的 Event Type，也就是動詞；(2) 改變 structured data 的 arguments，像是 Agent、Entity、Instrument。分別做完 (1)、(2) 後，我們可以得到兩個跟一開始的 event-centric structured data 差異不大，但是有些微錯誤的 structured data，接著同樣將這兩個 structured data 丟進 GPT，生出兩段文字(t1, t2)，而這兩段文字就是 negative samples。

將剛剛取得的三段文字 t0, t1, t2 做 Text Encoder，分別得到三個向量 t0', t1', t2'，將他們投影到 feature map 上，並且將原本的 image 做 Image Encoder，得到 v 向量，同樣投影到 feature map 上。而

Contrastive Learning 的目標就是讓 t0’ 與 v 在 feature map 上越接近越好， t1’、t2’ 與 v 越遠越好。

Q3：

main problem：先前只使用物體的 category 去進行學習，當遇到模糊不清的 category 的時候，模型往往會很難去辨識出這個 category 的 visual concept。

in this paper：

Step1：給定一個 category，然後在這篇 paper 中，設計了一個由 n 個 question templates 組成的 instruction T。我們要做的就是把 category 丟進 instruction T 裡面，作為 LLM 的 input，而 LLM 就會生成一段文字作為 output，我們稱這段文字為 Description。

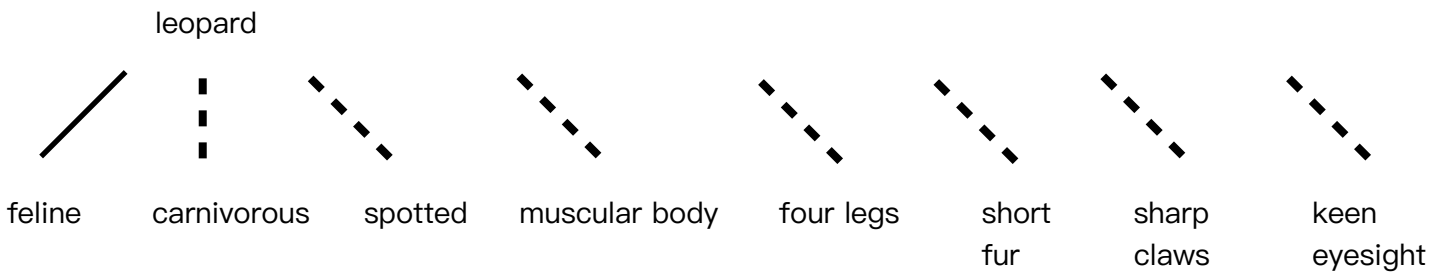
Step 2：額外設計了一個 instruction T’，接著我們將 T’ 跟在 Step 1 生成的 Description 結合起來，作為 input 丟進 LLM，而 LLM 就會生出我們想要的 structured graph- based data，其中包含了 Entity、Attribute、Entity-to-entity relation 以及 Entity-to-attribute relation。

Q4：

1. category 1： leopard

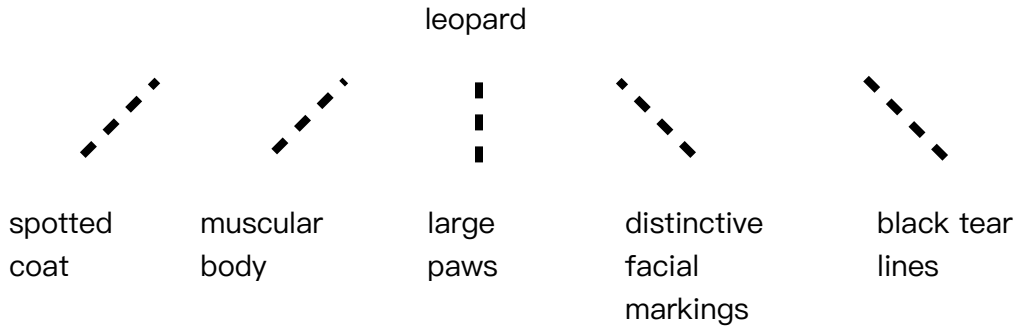
description 1："A leopard is a carnivorous, spotted feline with a muscular body, four legs, short fur, sharp claws, and keen eyesight.”

graph-based structured data component 1：



description 2："The distinct features of leopard are its spotted coat, muscular body, large paws, and distinctive facial markings including its black \"tear\" lines."

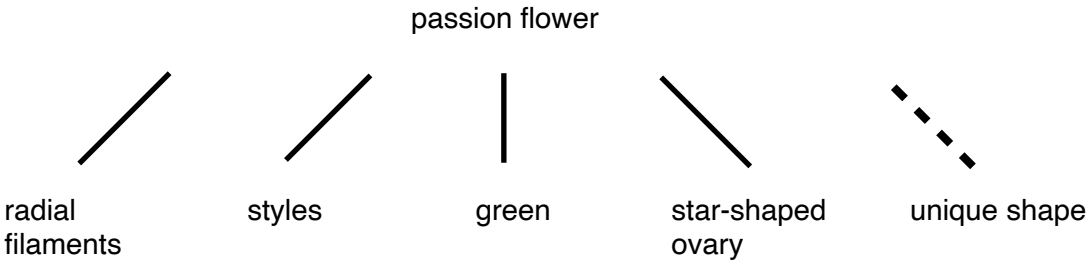
graph-based structured data component 2：



2. category 2： passion flower

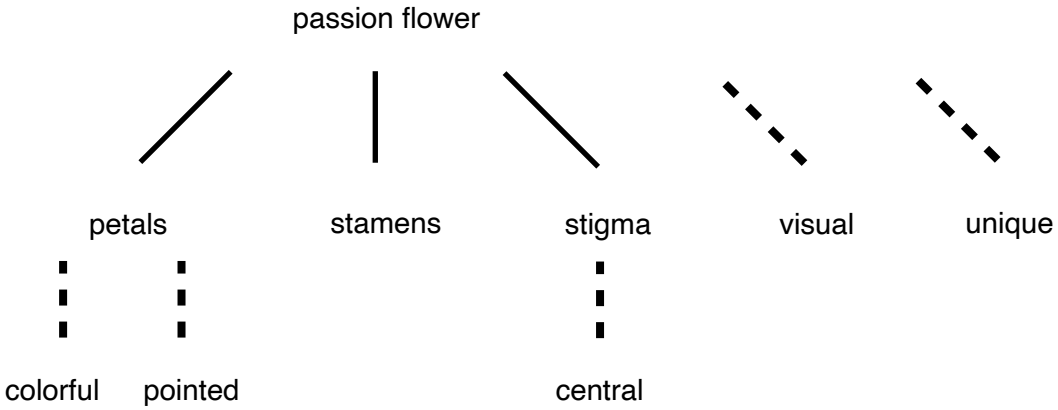
description 1："The passion flower has a unique shape with radial filaments and three styles that sit atop a green, star-shaped ovary."

graph-based structured data component 1：



description 2："The unique visual cue of a passionflower is its intricate arrangement of colorful, pointed petals and stamens enclosing a central stigma."

graph-based structured data component 2：



Q5：

我認為能夠將語言學的資料一併提供給 VLM 去進行學習。有了語言學的背景知識，model 在與使用者進行對談、處理問題時，更能夠了解使用者每段文字背後欲表達的內容。即便有些時候使用者的問題並不明確，但 model 藉由語言學的知識，推論使用者這句話背後的真正意圖，進而讓 model 可以更精準的抓到使用者的問題點，給予更正確、明瞭的回覆。同時，我們還能夠將一些抽象的特性，例如愛、開心等情緒配合著相對應的圖片庫給 VLM 進行學習，或許就有機會讓這個 model 透過一張圖片，就能夠獲取圖片中的人當時的情緒，從而做進一步的分析。有了越多的 depp semantic knowledge，model 距離成為 AGI 或許又更近了一步，但將如何將這些複雜、抽象的資訊有效地整合在一個大型模型裡，將會是人類將面臨的一個重大挑戰。