AI 實驗室

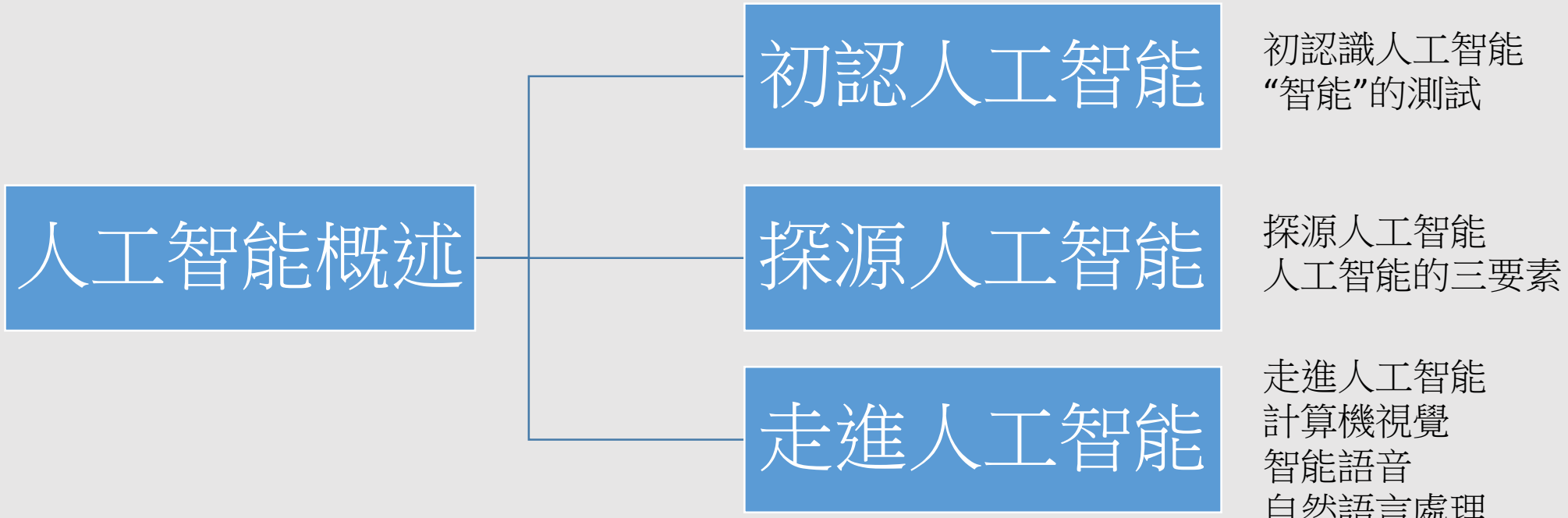
Edit by Hammond Lai

內容參照”走進人工智能”

目錄

- 人工智能概述
- 會學習的人工智能
- 會看的人工智能
- 會聽說的人工智能

(第一章) 人工智能概述


人工智能概述
初認人工智能
探源人工智能
走進人工智能
初認識人工智能
“智能”的測試
探源人工智能
人工智能的三要素
走進人工智能
計算機視覺
智能語音
自然語言處理

初認識人工智能

- 智能是什麼？迄今爲止，人類對於這個問題並沒有廣泛的共識。人類運用思維解決問題。這其中包含邏輯推理等一系列動作。但事實上，人類解決問題的方法往往並不唯一，甚至有時候我們很難判定誰的方法更好。智商測試儀器也許能測試一個人的智商。但是一定能碩智商高的人擁有更多智能。對社會更有貢獻嗎?

- 答案並非如此。首先，高智商不行代表高智能，也代表著創新能力。善於思考，具有發現問題、解決問題的能力。其次，除了智商，我們還需要有一定的情商。而情商是認識自我、控制情緒、鼓勵自己以及處理人際關系、參與團隊合作等相關的個人能力的總稱。最後，我們還需要有正確的價值觀。能分辨是非。具有甄別能力。綜上所述。也許，這才是人類社會更需要的智能 .

- 對於人工智能的定義，目前較爲普遍的認知是，人工智能是研究、開發用於模擬、延伸和擴展人的智能理論、方法、技術及應用系統的一門新的技術科學 。

“智能”的測試

圖靈*認爲機器是可以進行思考的，但這需要機器具有快速的運算速度，具備超出人類的邏輯單元與記憶容量，同時編輯編制大量的智能化程序。並提供恰當類別、足夠數量的數據。如何判斷機器是否具備思維能力呢？
在圖靈測試提供了一種檢測機器是否具備人類智能的方法。

簡單來說，就是將一 台 機器A和一個人B作爲被測試者。再選一個人C作爲測試者。這裏需要將測試者和被測試者隔離。也就是將機器A和人B放在一個房間，人C放在另一個房間。

通過一些裝置(如鍵盤)，人C向機器A和人B進行隨意的提問，經過幾輪提問和回答後。如果有30%的回答讓作爲測試者的人C，無法區分是由人的還是機器A做出的，那麼就可以認爲這台機器A具有智能 。

* 艾倫圖靈是被譽爲”計算機科學之父”和”人工智能之父”的英國數學家，邏輯學家、密碼學家。計算機領域的最高獎項，圖靈獎就是以其名字命名的。

探源人工智能

- 人工智能的第一次浪潮，20世紀50至80年代。

  Shakey移動機器人，通過解覺感知、運動、規劃與控制問題，能夠尋找目標，並把它移動到指定的方位，實現簡單的自主導航。 Shakey 移動機器人印證了很多屬於人工智能範疇的科學結論

  Eliza是能與人對話的人工智能程序。它並非依靠聲音，而是利用文本與人進行對話，基於對話規則與模式匹配，根據人類語句中的關鍵詞給予相應的回答

- 人工智能的第二次浪潮，20世紀80至90年代末期。

  專家系統就是關注於某一些特定領域的系統。從專門智識中推演出規則。模擬專家解決問題的流程，利用知識得到滿意的解答。例如，醫療上爲特定患者進行診斷，看其是否患有特定的疾病專家系統往往不是通用的，而是爲解決某個問題，或者爲實現某個具體目標而進行發開發的

- 人工智能的第三次浪潮，21世紀初至今。

人工智能的三要素

- 當今人工智能的發展離不開**數據**、**算法**、**算力**，因此數據、算法、算力被稱爲人工智能的三要素。

數據在我們的生活中充滿着各種各樣的數據，坐地鐵留下了進出站的信息。去圖書館借書，會留下借書的信息，在學校學習會留下電子成長檔案，以上的信息都是數據。

算法是解決問題。實現目的的方法。可以理解爲，爲了解決某個問題，先思考怎麼做，用什麼方法在採取行動，進而解決問題。算法具備以下特點。

1. 有窮性　生物的生活。算法必須能在執行有限個步驟之後終止。
2. 確定性　算法的每一個步驟必須有確實的含義。
3. 可行性　算法中每一個步驟都是要能夠實際做到的，而且是在有限的時間內完成。

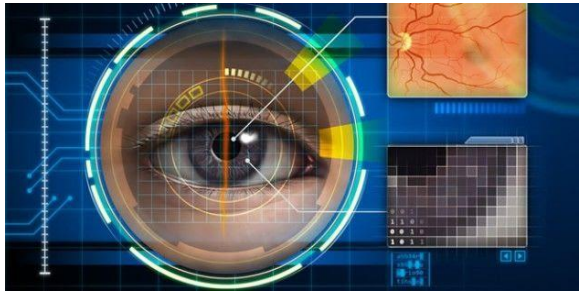
算力通常表示計算機的計算能力。算力的提升可以認爲是個系統工程，不涉及及諸如芯片、內存、硬盤等所有硬件組件，而且要根據數據類型的應用所處的實際環境對計算架構，對資源的管理和分配進行優化。

走進人工智能

在許多電影中。機器人都具備人工智能，但是我們需要知道，機器人僅僅是人工智能的一種載體。並不是說人工智能就等於機器人。

人工智能包含許多技術，

機器能像人一樣會"看"，是使用了計算機視覺技術。



會"聽"使用了智能語音技術中的語音識別。

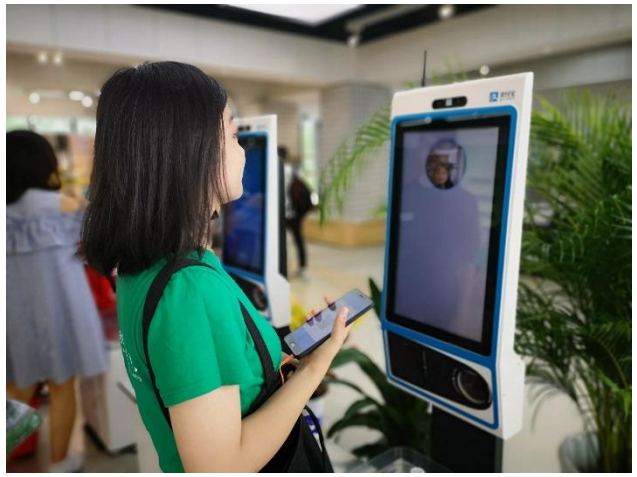
會"說"是使用了智能語音技術中的語音合成。

會"學習"是使用了機器學習技術

計算機視覺 (電腦視覺)

明星演唱會有時還附加了”抓捕逃犯”的功能。僅通過十華語樂壇一位明星在2018年的巡回演唱會，前前後後共抓捕55名在逃犯人 。 這些抓捕行動得以圓滿完成，背後有一個大工程，天網工程。天網工程的核心技術之一就是計算機視覺的圖像處理與識別技術，通過設計在現場的高清攝像頭，清晰地捕捉到演唱會現場觀眾的面部特徵，並且迅速和公安系統信息庫中的存儲的信息進行對比，一經發現疑犯，立即採取行動抓捕，真正做到了天網恢恢，疏而不漏 。

計算機視覺除了可以幫助抓捕逃犯，在日常生活中也給我們帶來了很多便利。例如手機上的人臉識別功能，我們經常用來解 鎖手機、完成支付等，人臉識別考勤系統在很多公司也得到了應用。



Edit by Hammond Lai

智能語音

在人機語言對話中，智能語音技術是必不可少的。智能語音技術包含語音識別技術與語音合成技術，前者將語音轉成文字。後者將文字轉爲語音，智能語音技術賦予機器能聽會說的本領。在日常生活中，我們使用語音輸入法進行文字輸入，使用手機導航時聽到語音播報，這些都離不開智能語音技術。除了語音、智能語音技術以外，爲了能讓機器更好地理解人類的語言，說話更符合人類的說話方式，自然語言處理技術在其中起了積極作用。自然語言處理技術與智能語音技術雙劍合璧，使得人類與機器進行語音對話更順暢。

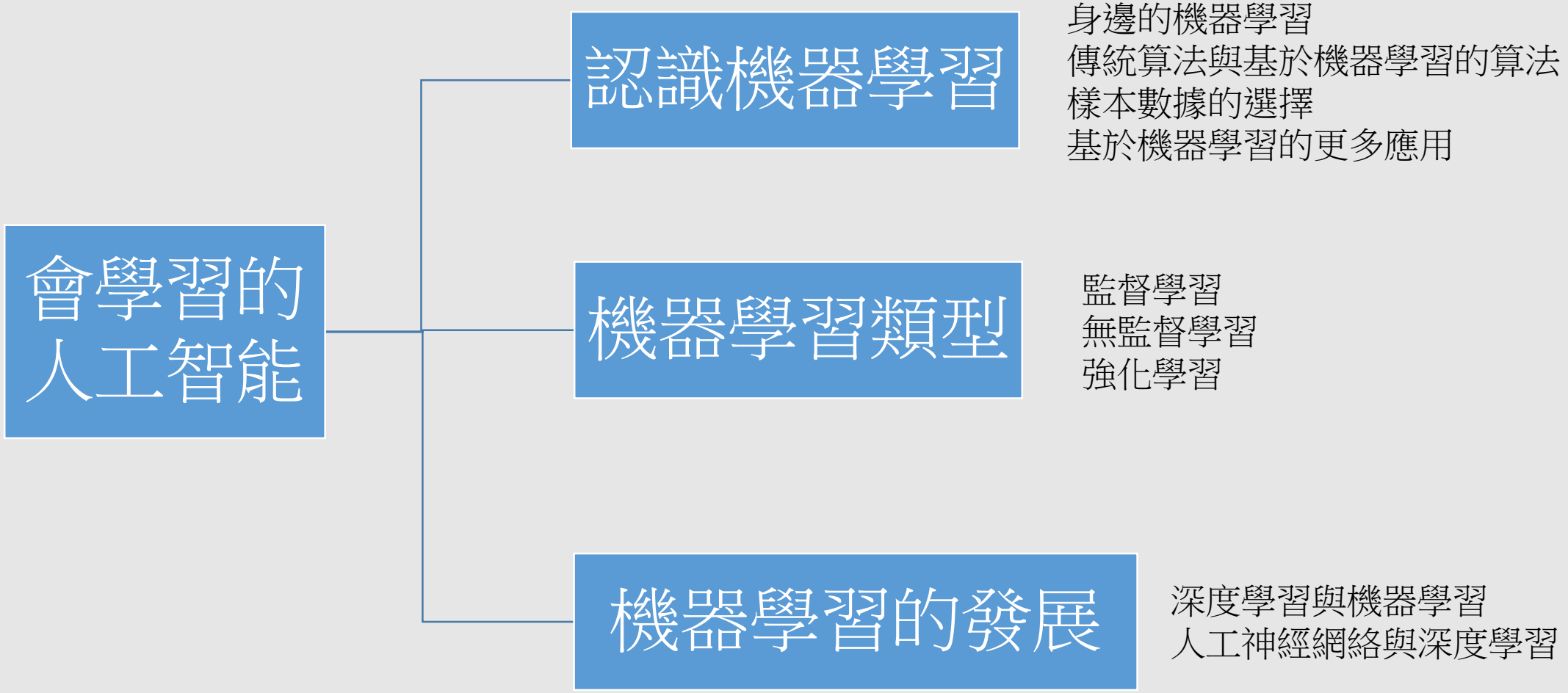
自然語言處理

- 語言是人類區別於其他動物的特徵。人的許多智能都和語言有密不可分的聯系，自然語言是人類日常使用的語言。

- 機器翻譯屬於自然語言處理其中的一種。世界上有那麼多的種族，各種族大都有自己的語言體系。各國語言不通阻礙了大家的交流，是一個令人頭疼的問題。作爲一個普通人，想玩成”世界那麼大，我想去看看”的願望，語言關必須要過，但是學好一門外語已經不容易了，更何況還有那麼多的語言，普通人根本無法全部掌握。

- 隨着翻譯機的出現，現在這個問題已經得到了很好的解決。以我國強大先進的翻譯機爲例，它不僅覆蓋了。全球近200個國家和地區的語言，而且還能識別出加拿大、英國、澳大利亞、印度，新西蘭五個國家的帶口音的英語和我國的四種方言(粵語、四川話、東北話、河南話)。這個功能非常了不起，並且非常實用。

(第二章)會學習的人工智能



Edit by Hammond Lai

身邊的機器學習

2016年一場人機圍棋大戰。阿爾法圍棋(AlphaGo)以4：1的總分戰勝了圍棋世界冠軍李世石，人們都在問，阿爾法圍棋是什麼？他是怎麼做到的？
AlphaGo 圍棋的核心技術是機器學習，機器學習的思想其實並不復雜，它是模擬人類生活中的學習過程，機器學習依賴於數據，數據是承載着信息的各種符號組合。
AlphaGo Zero 用三天時間完成了490 萬局自我對弈。 幾天內它就發展出擊敗人類頂尖棋手的技能，而早期的 AlphaGo要數月的訓練才到達同樣水平

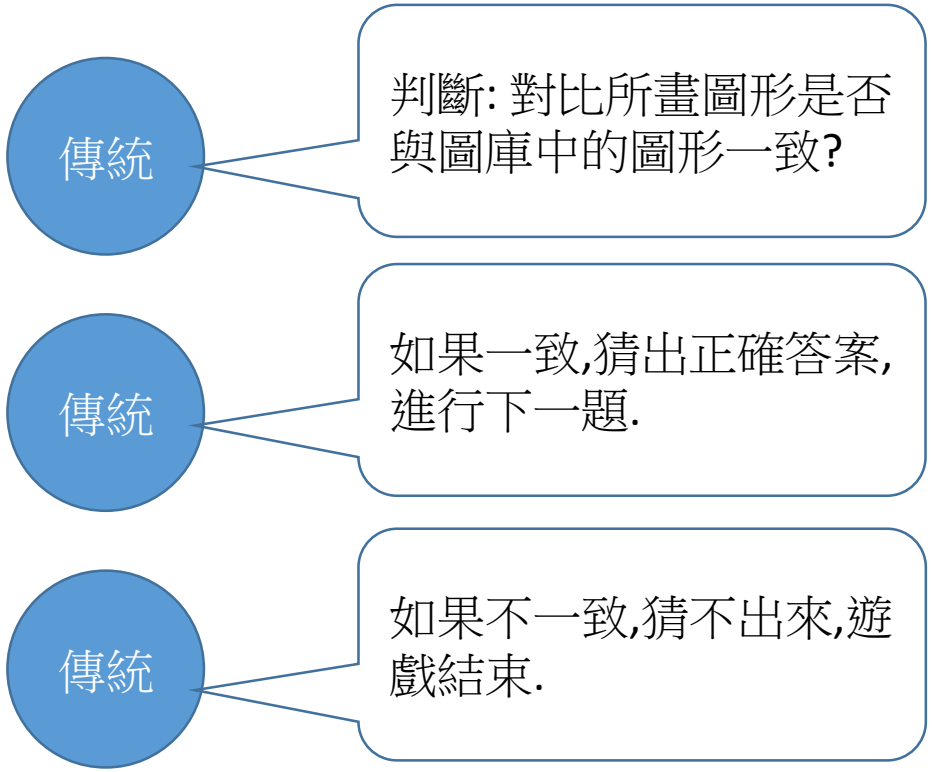
數據不僅是狹義上的數字，還可以是文字、符號、圖片、視頻、音頻等。數據隨時隨地都在產生，比如在社交平台上發布的各類信息，在購物平臺上購買物品的記錄和發布的評論等都是數據。這些數據產生後會被採集、處理和加工，從而轉換成可供計算、分析和使用的新數據。

https://www.youtube.com/watch?v=tXlM99xPQC8

傳統算法與基於機器學習的算法

- 傳統算法

傳統算法是編寫具體識別規則的程序，程序的判定過程如圖所示。在傳統算法編寫的程序中。預存許多圖片數據作爲圖庫，這些圖庫也稱爲樣本數據。傳統算法需要對比玩家所畫圖形是否與圖庫中的圖形一致。若圖形不在圖庫中，就會導致猜不出來，因爲程序在執行過程中是完全遵守所編寫指令運行的。事實上，每個人畫畫的水平不同，畫風不同，要畫的與圖庫一致是非常困難的



傳統算法與基於機器學習的算法

- 基於機器學習的算法

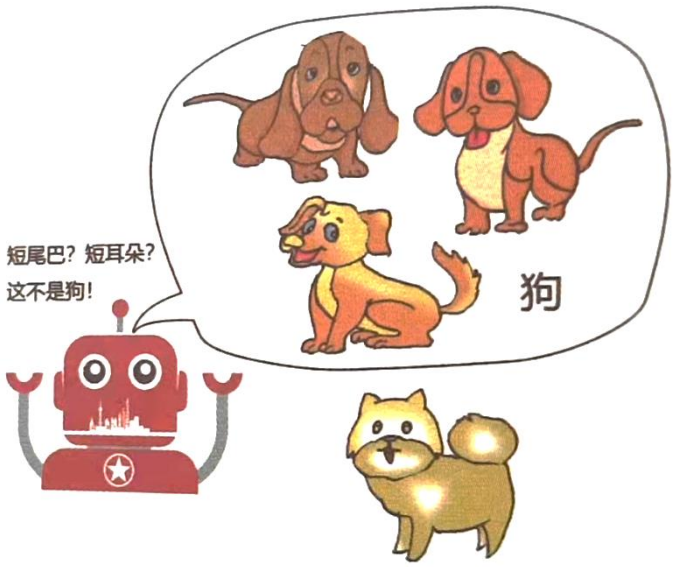
“猜畫小歌”使用的是基於機器學習的算法，在這個過程中，他並不編寫具體的規則，而是讓機器從 “猜畫小歌”圖庫數據庫裡自動學習集中的規律。這個過程可以理解爲機器自動從樣本數據中提取特徵，再將具有相同特徵的圖形進行歸類，然後對每一類建立模型之後，判定用戶畫的圖更接近那哪一類模型。比如。當我們畫一只狗。“猜畫小歌” 通過比對特徵點，最終將其分到 “猜畫小歌”的狗那一類。從而猜出所畫的是一只狗。這一機器學習流程如右圖所示.

樣本數據的選擇

- 想要通過分類器學習並建立模型是需要一定數據量的.採集的樣本

  (1)數據量過少,達不到學習所需的數據量.

  (2)或者採集的樣本數據量雖然很多，但是樣本數據種類過於單一，都會導致分離器出現誤差



比如 “猜畫小歌” 想要分辨玩家所畫的是不是狗，如果圖庫中的樣本數據是存儲了幾張長耳朵長尾巴的狗，由於樣本數量太少，當玩家畫了松鼠狗這類短耳。短尾巴的狗時，“猜畫小歌” 就無法判定這是狗了。

比如狗的樣本，數據量雖然很多，但都是松鼠狗這樣的卡通形象。“猜畫小歌” 會把松鼠狗的所有特徵都記住，此時如果玩家花了一個法國短腳獵犬的卡通形象，那麼 “猜畫小歌” 有可能認爲法國短腳獵犬和松鼠狗的區別太大，從而判定巴吉度獵犬不是一只狗

基於機器學習算法的更多應用

- 搜索引擎中的機器學習

  早期的搜索引擎基本是對高質量網站的羅列，後來才發展成自動索引技術。垃圾垃

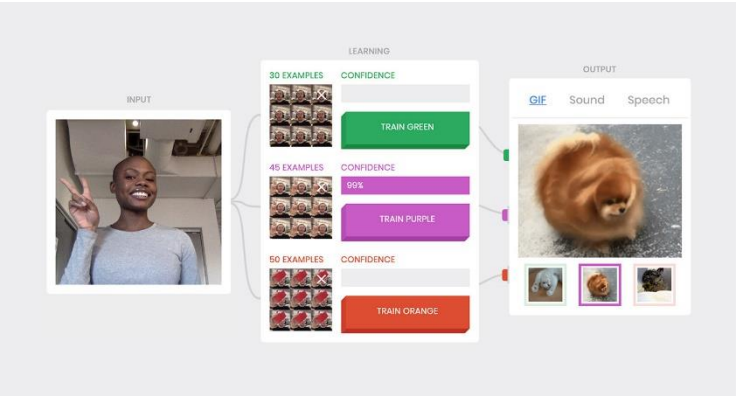
- 垃郵件過濾器中的機器學習

  使用電子郵件系統時，垃圾郵件過濾器能幫助我們過濾大量的垃圾郵件。也是因爲運用了基於機器學習的算法，爲了確定垃圾郵件過濾器是不斷更新的，他們使用的機器學習算法讓機器不斷提取垃圾郵件的特徵，幫助用戶分類這些垃圾郵件

- 人物分類中的機器學習

  是用手機照片中人物分類功能時，它能識別照片中的人物，並能找到有這個人的所有照片，這也是使用的基於機器學習的算法

除了上面分享的應用場景外，生活中還有很多用到機器學習的地方。基於機器學習的算法也在人工智能許多技術中得到應用。比如語音識別、人臉識別等，機器學習，可以說是已經融入人們生活中的方方面面



Edit by Hammond Lai

監督學習

- 生活中，我們通過很多方式來學習。在我們很小的時候。爸爸媽媽會拿着蘋果告訴我們，像這樣紅色的、球形的，表面光滑的就是蘋果。
- 後來，當我們看到了青蘋果時，爸爸媽媽會告訴我們，其實蘋果都不是紅色的，也有青色的。漸漸地我們在腦子裏面有了蘋果的模型，後來我們就能分辨出蘋果了。
- 等我們長大了，會自己看書，也能從書中自學，找到事物之間的關聯，有時我們也會通過不斷試錯來學習。

- 機器學習也有自己的學習方式，如監督學習、無監督學習、強化學習等

監督學習

人類的學習過程去認識美西螈:

1. 看
2. 記憶
3. 在大腦中形成模型
4. 驗證模型

外型		顏色
腮	有六個像角一樣的腮	顏色有白色的,黃色的,黑色的.
臉	臉寬大	
腿	有四條腿並且腿比較短,比較細	
尾	有著魚一樣的尾巴	



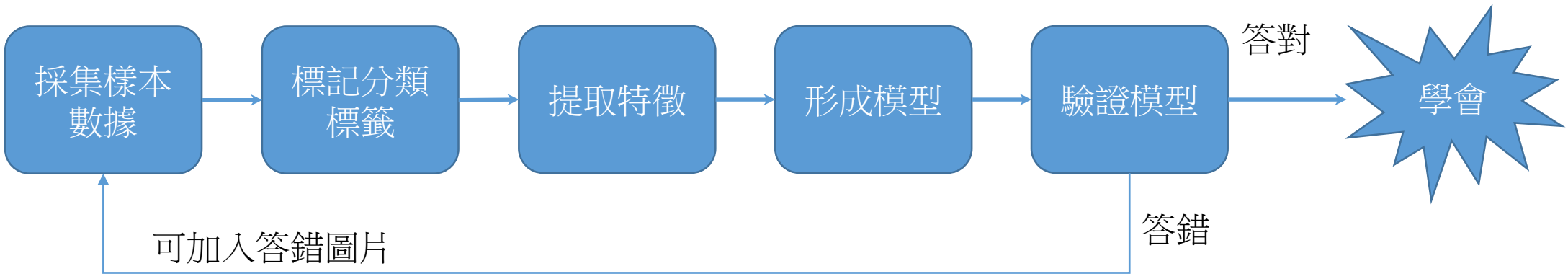


監督學習的一般過程

- 像人類一樣通過觀察已經被告知內容的圖片進行學習，這些都是”美西螈”的照片的方式。在機器學習領域被稱爲監督學習。監督學習通常表示機器學習的數據是帶有標記的，比如通過監督學習的方式認識”美西螈”時，需要提供給機器標記爲 ”美西螈”的圖片。在這種學習過程中，標記(Features)就是人給予機器的引導。監督學習的優點是學習效率高，但是他也有明顯的缺點。過於依賴人，當人的引導出現失誤。
- 比如當機器進行監督學習時，在人給的大量帶有標記(Features)的樣本數據中，夾染一些不是 ”美西螈”的圖片數據，那麼機器就可能不會準確地提煉出 ”美西螈”的特徵。由於這樣的機器學習方式非常依賴人的知識面，所以在一定程度上限制了機器的學習範圍。例如，如果人也不認識 ”美西螈”，那麼就沒發給出帶有標記(Features)的樣本機數據。機器在監督學習的方式下，就沒法進行”美西螈”的學習，也就不會認識 ”美西螈”.

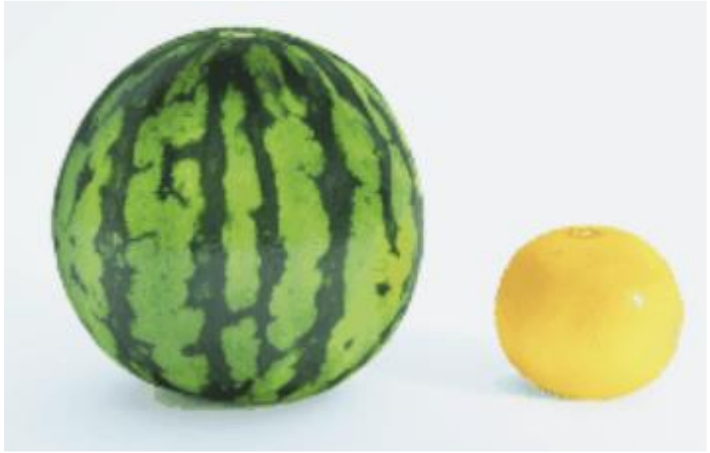
監督學習的一般過程

1. 採集帶有標記的樣本數據
2. 提取特徵
3. 形成模型
4. 驗證模型



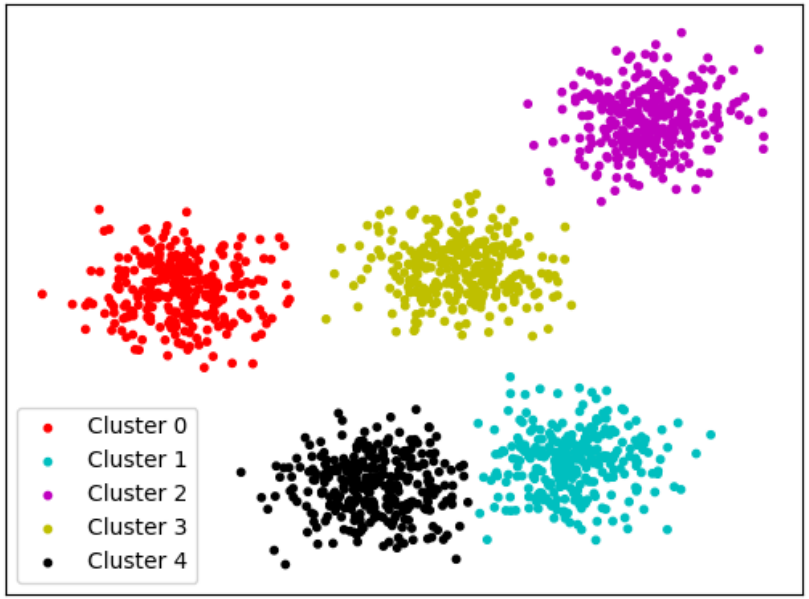
無監督學習

- 我們看到蘋果、山楂、草莓、红辣椒時，可以根據它們的顏色進行歸納，他們共同的顏色是紅色。又比如我們看到足球、橙子、西瓜時，可以根據形狀進行歸納，他們共同的形狀是球體，這種學習方式就是歸納式推理學習。人類自行發現相同點和不同點，根據這些特點分成不同的類別。並且給這些類別取名字，比如”紅色”，”球體”。



無監督學習的一般過程

- 人類這種歸納推理是學習。對應到機器學領域。就是無監督學習。在這種學習過程中，沒有人引導和監督，也就是輸入沒有標記(Features)的數據。它是一種機器自主學習的方式。正因爲有了無監督學習，機器開始自己探索未知的世界。無監督學習常用的方法是"聚類"， 所謂類，通俗地說，就是指相似元素的集合。
- 機器自主的將數據分隔爲若幹個小個體。這樣的小團體也可以稱爲"集羣"，需要注意，雖然聚類在無監督學習中發揮了積極的作用，但是並不意味着"無監督學習等於聚類"。



無監督學習的一般過程

- 今天去動物園看到老虎和熊貓。
- 老虎和熊貓的觀衆最多。
- 老虎是食肉動物。
- 午飯吃蛋炒飯和羅宋湯。
- 蛋炒飯裡還有香腸。
- 明天還想吃蛋炒飯和羅宋湯。

把最相似的集羣合並，歸納成一個新的集羣。可把語句中包含相同名詞最多的定義爲最相似。這裏將各句話中包含的相同名詞數看作相似度

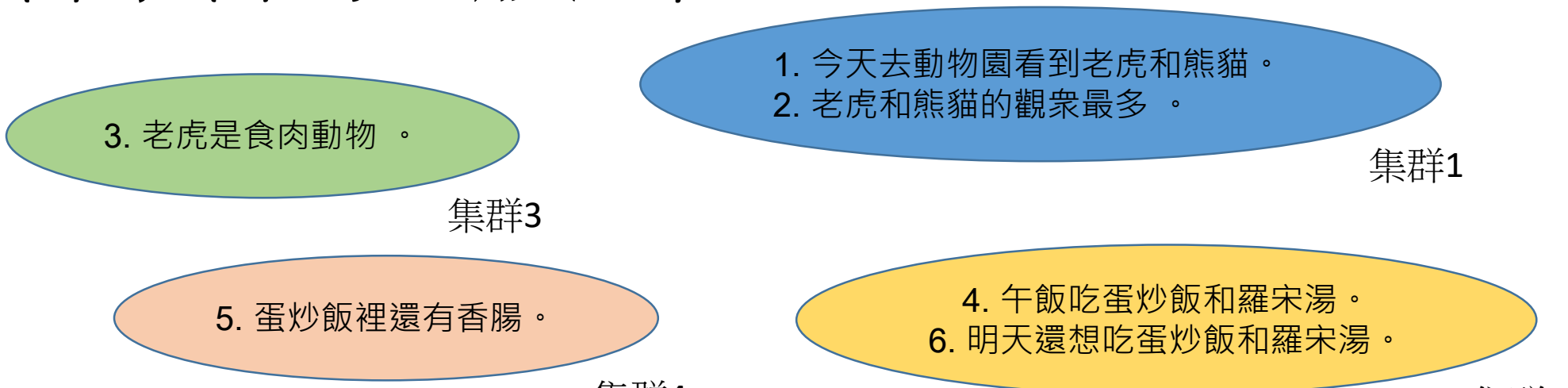
標記各語句中的名詞

句子	老虎	熊猫	蛋炒饭	罗宋汤	香肠
①	√	√			
②	√	√			
③	√				
④			√	√	
⑤			√		√
⑥			√	√	

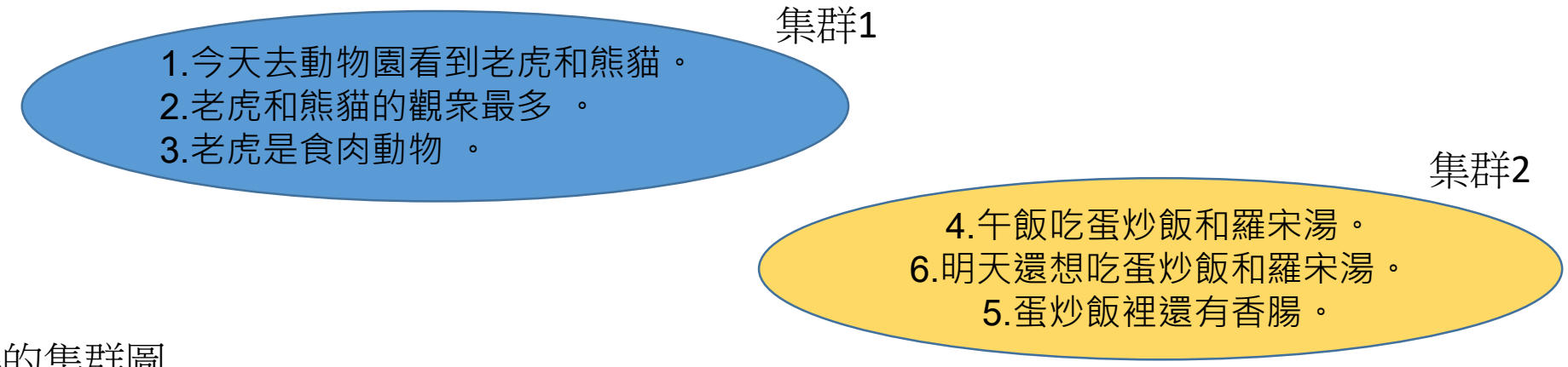
各語句對應其他語句的相似度

句子	①	②	③	④	⑤	⑥
①		2	1	0	0	0
②	2		1	0	0	0
③	1	1		0	0	0
④	0	0	0		1	2
⑤	0	0	0	1		1
⑥	0	0	0	2	1	

無監督學習的一般過程



相似度為2 時新獲得的集群圖



相似度為1 時新獲得的集群圖

監督學習與無監督學習的比較

	監督學習	無監督學習
有分類標籤與母分類標籤	監督學習是輸入帶有分類標籤的數據	無監督學習是輸入沒有分類標籤的數據
有確定輸出結果與無確定輸出結果	監督學習的輸出結果就是預設的分類標籤，因此該結果是確定的	無監督學習的輸出結果是事先未知的，在不同算法，不同相似情況下輸出的類別結果是不確定的。比如我們觀察聚類後的結果，他可能按顏色進行聚類，也可能按形狀進行聚類
分類與聚類	分類是監督學習的一種方法	聚類是無監督學習的一種方法

強化學習

- 人類用思考與改進的方法提升自己對知識的掌握和認知。在人工智能機器學習中，我們將這個過程稱爲強化學習
- 訓練導盲犬例子就是人類思考與改進的例子. 導盲犬是怎麼做到的呢？

  首先，他會觀察周邊的環境，包括障礙物的位置、障礙物的多少等，思考前進的路線，然後開始進行行動。當他能夠帶領訓練員安全通過障礙物時，會得到表揚，會被撫摸。這讓他很開心。但是，當他沒能安全帶領訓練員安全通過障礙物時，會被訓練員批評，這讓他很失落。爲了不被批評，並且獲得更多的表揚和撫摸，它就會思考改進避讓障礙物的路線。不斷鞏固成功的線路，改變失敗的路線。隨着訓練次數的增加，不斷的循環，倒忙全就能夠成功上崗，帶領訓練員通過各種障礙物。

Edit by Hammond Lai

強化學習的一般過程

- 強化學習可以看作機器思考與改進的過程。
- 強化學習其實是一種不斷試錯的學習，在各種環境下，需要盡量嘗試所有可以選擇的動作，通過環境給出的反饋 (即獎懲)來判斷動作的優劣，最終獲得環境和最優動作的影射關系(即決策)。
- 選擇一條線路，從起點到終點，經過每個節點陷入能夠獲得一定數量的金幣(藍字表示每個節點路線的金幣數)，那麼選擇哪一條鮮鹿獲得金幣的數量最多呢？

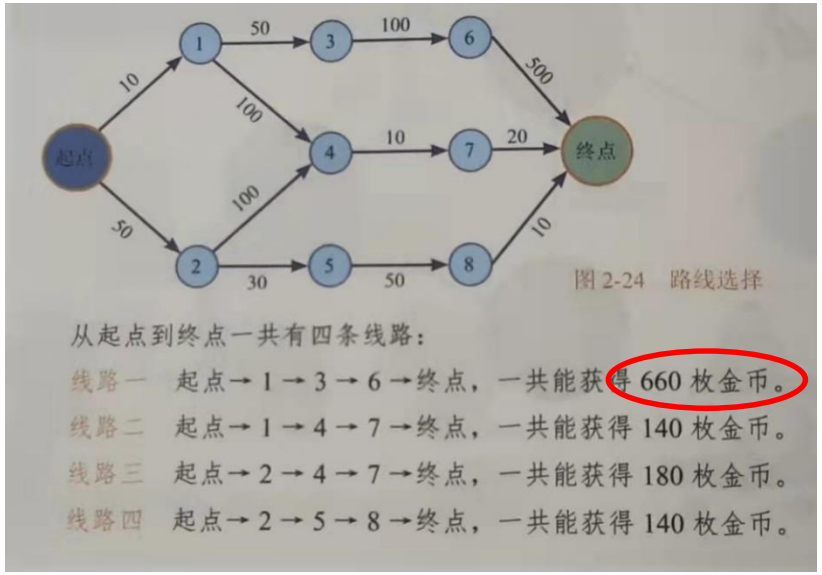


图 2-24 路线选择

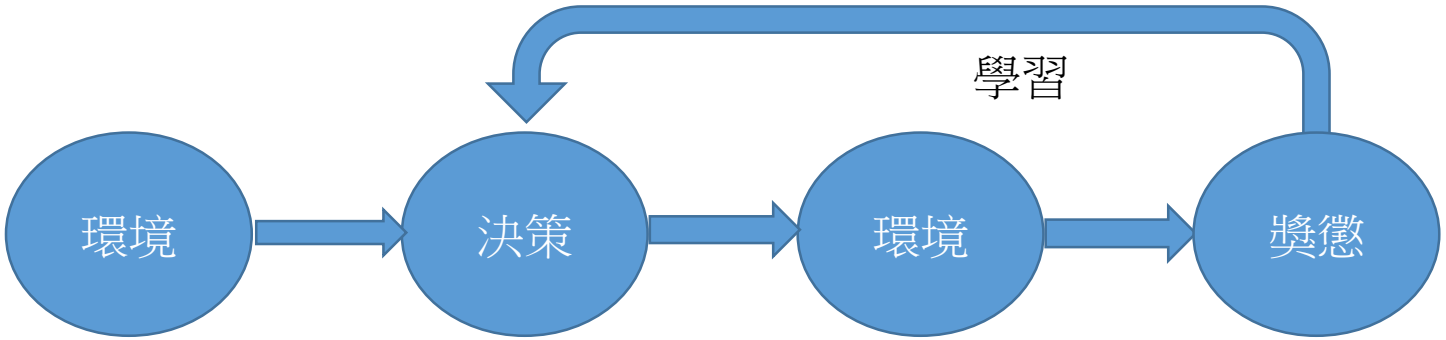
从起点到终点一共有四条线路：

线路一 起点→1→3→6→终点，一共能获得 660 枚金币。

线路二 起点→1→4→7→终点，一共能获得 140 枚金币。

线路三 起点→2→4→7→终点，一共能获得 180 枚金币。

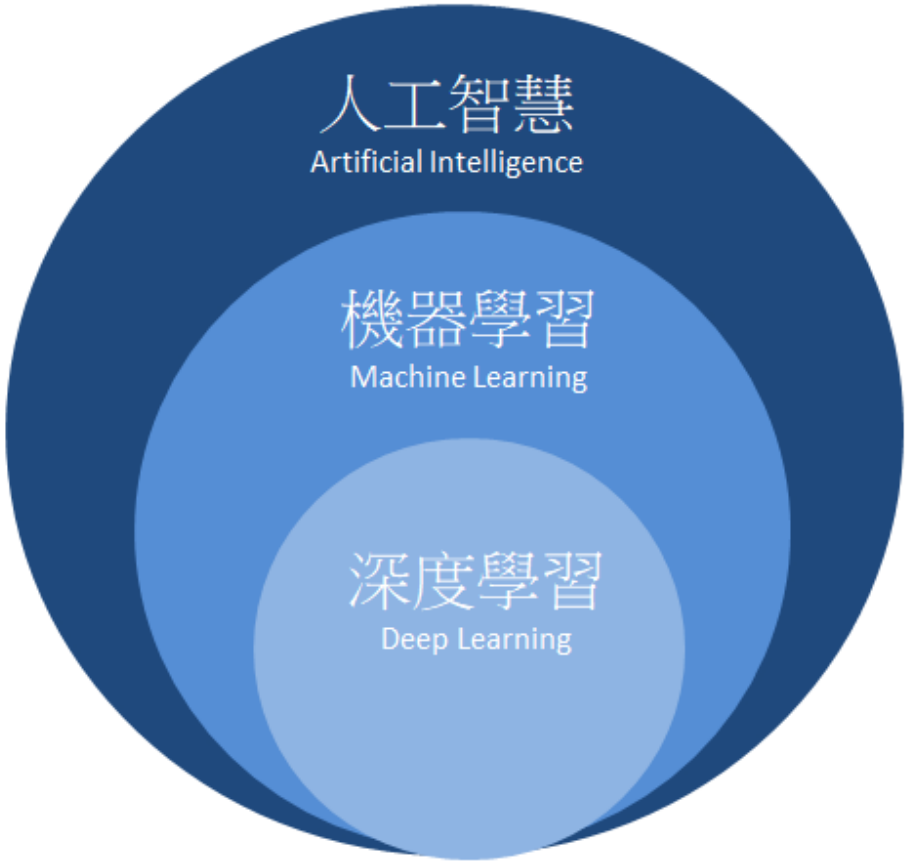
线路四 起点→2→5→8→终点，一共能获得 140 枚金币。



Edit by Hammond Lai

深度學習與機器學習

人工智能中的智能可以說主要歸功於機器學習，目前在機器學習中經常使用的技術就是深度學習。簡而言之，我們之前提到的”監督學習”、”無監督學習”、”強化學習”這三種學習方式都能使用”深度學習”技術

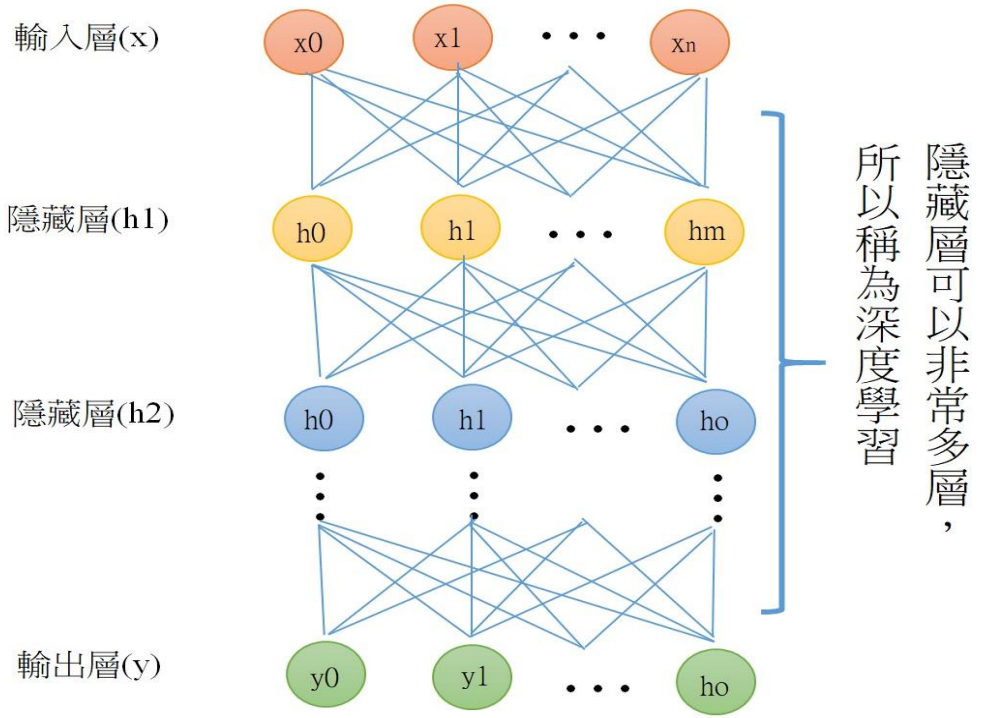


人工神經網絡與深度學習

- 神經元
  人類的大腦中擁有數以千億計的神經元，神經元是構成大腦的基本部件。
- 感知機
  感知機是神經網絡的基礎，也是深度學習的起源。單層感知機就可以看作一個最簡單的神經網絡。
- 多層感知機

  單層感知機過於簡單，能處理問題有限，因此多層感知機應運而生。多層感知機就是在輸入和輸出中增加更多層，並且輸出層也不僅僅是一個神經元，可以爲多個神經元。
- 人工神經網絡

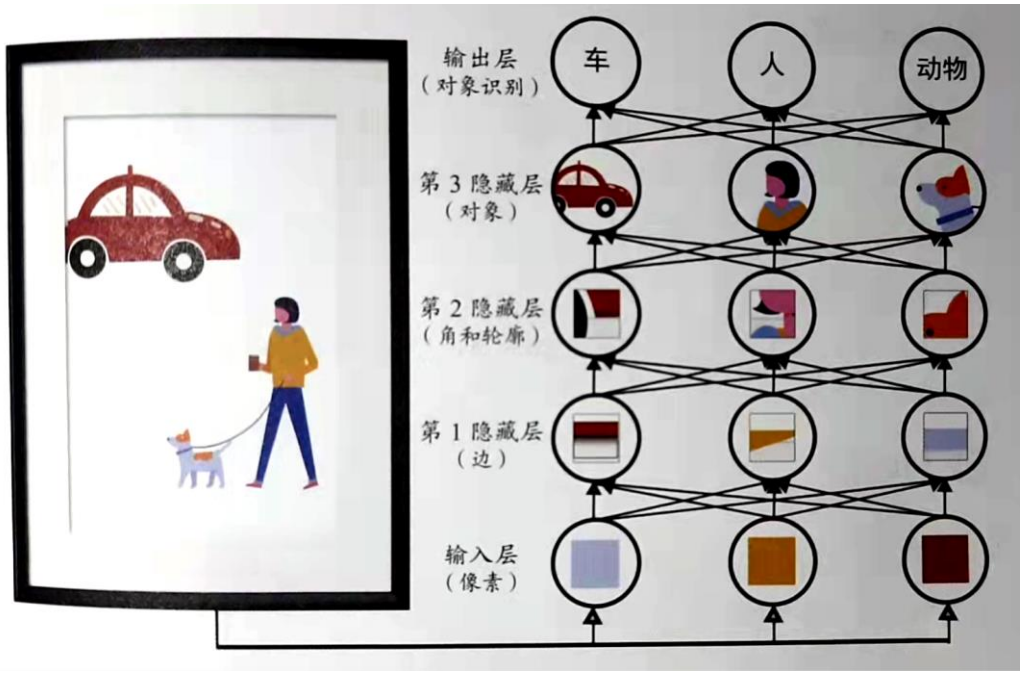
  我們可以將人工神經網絡看作高度模擬大腦結構的成果。人工神經網絡分爲輸入層、隱藏層、輸出層。輸入層負責數據的輸入，中間幾層神經元數據由於無法看到，稱爲隱藏層，輸出層負責數據輸出。



深度學習

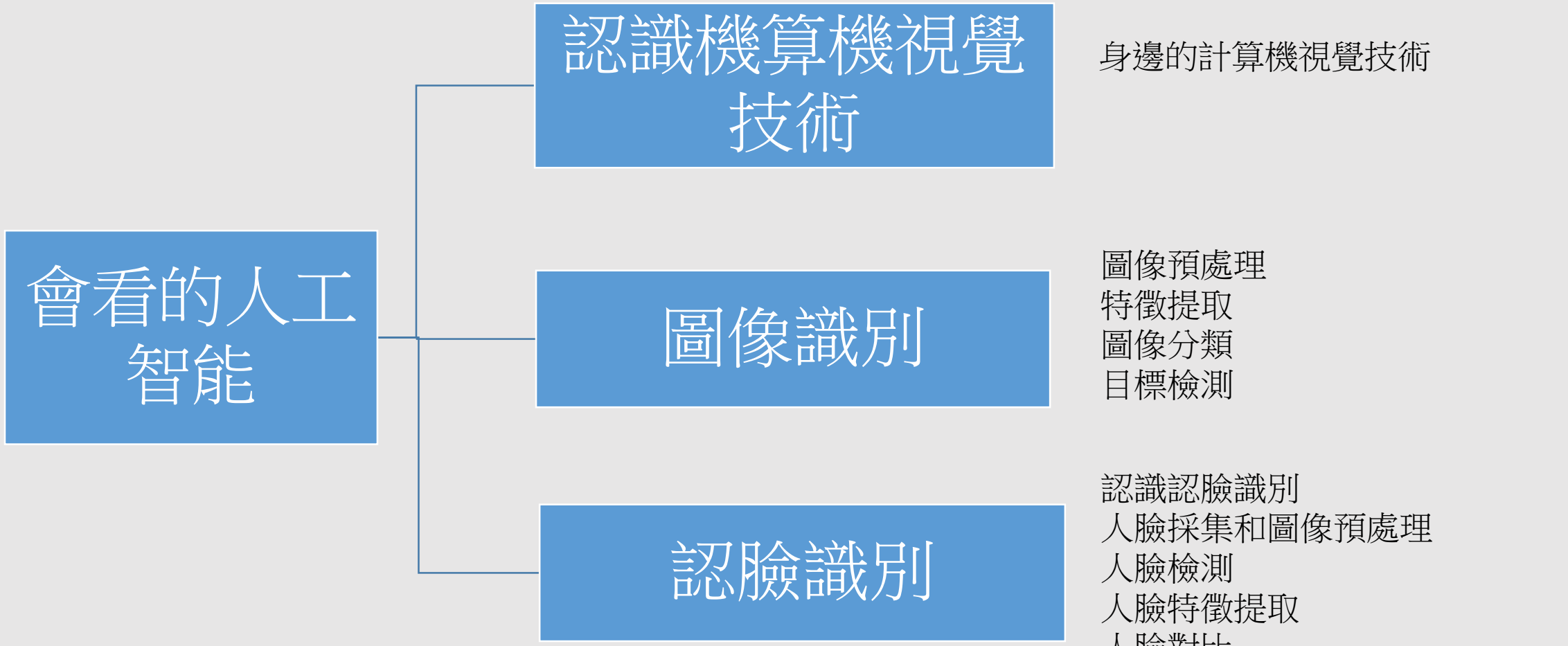
- 每一張圖片都是有一個個像素構成的。我們將像素作爲輸入層的輸入數據

- 隱藏層負責從圖像提取更多的抽象特徵
- 根據給定像素。第一隱藏層，通過比較相鄰像素的亮度來判定邊緣。
- 有了第一層隱層層的邊緣描述，第二隱藏層通過搜索邊緣判定角和輪廓。
- 有了第二隱藏層中的關於角和零闊的描述，第三層隱藏層會通過搜索圖像中的角和輪廓可以檢測特定對象的整個部分
- 最後，根據圖象所描述的對象進行識別，從而通過輸出層輸出。



- 當然，根據設置的深度學習模型不同，隱藏層的層數與每一層神經元的個數也會不同，每一個隱藏層的功能也會不同。比如換一個深度學習模型，其地隱藏層可能並不是在判定邊緣

(第三章) 會看的人工智能


會看的人工智能
認識機算機視覺技術
圖像識別
認臉識別
身邊的計算機視覺技術
圖像預處理
特徵提取
圖像分類
目標檢測
認識認臉識別
人臉採集和圖像預處理
人臉檢測
人臉特徵提取
人臉對比
計算機視覺技術更大範圍的應用

身邊的計算機視覺技術

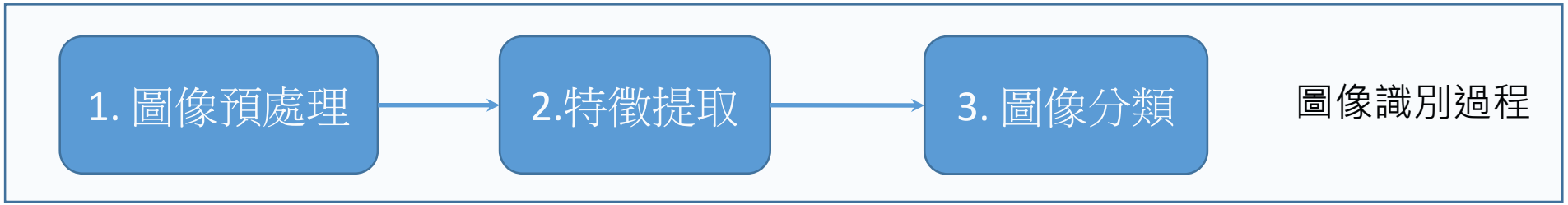
- 我們學習和生活中會遇到各式各樣的攝像頭:
- 學校門口用來測量體溫的攝像頭，教學大樓裏安防攝像頭，在教室裏用來點名的攝像頭。教室的攝像頭，甚至可以在課堂互動時監測到閱讀，舉手、手寫、起立、聽講、趴桌子等行爲，再結合面部表情，是高興、傷心或是憤怒、反感，分析出在我們在課堂上的狀態。

- 在上下學路上，我們可以看到地鐵、公共汽車裏和馬路上的監控攝像頭，汽車進出小區時欄杆上遊可以識別車牌號碼的攝像頭。

圖像識別

- 圖像識別，是指利用計算機對圖像進行處理、分析和理解，以識別各種不同模式的目標和對象的技術。現階段，圖像識別技術，從識別對象上一般分爲人臉識別與物品識別。

- 人臉識別主要運用在安全檢查、身份檢驗與移動支付等場景中，物品識別主要運用在物品流通等場景中，特別是無人貨架、智能零售櫃等



圖像識別

图 3-18　原图　　图 3-19　灰度校正后　　图 3-20　几何变换后

1. 圖像預處理

- 圖像質量的好壞直接影響識別算法的設計與識別效果的精度。因此，在圖像識別的初始階段，我們需要對圖片進行預處理。圖像預處理的主要目的是消除圖像中的誤差，恢復有用信息，從而改進特徵提取、圖像分類的可靠性，幫助計算機更準確地理解圖片
    - 灰度校正

      我們把圖片上每一個不可分割的單位(或者說是最小的元素)稱爲像素"灰度"一詞。原始黑白攝影的術語，是指根據景物各點顏色及亮度不同。攝制成的黑白照片(或黑白圖像)上的各像素點呈現出不同深度的灰色。
    - 幾何變換

      圖像幾何變換又稱爲圖像空間變換，是指通過平、移轉、置鏡像、旋轉、縮放等方法對採取的圖像進行處理。可以比較好的糾正圖像採集中圖像太暗、圖像太亮，有噪聲點、對比度不明顯、成像角度等問題。

圖像識別

图 3-21　原图

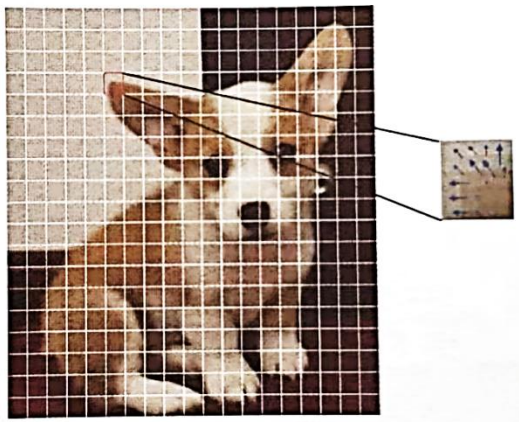


图 3-22　梯度变化示意图

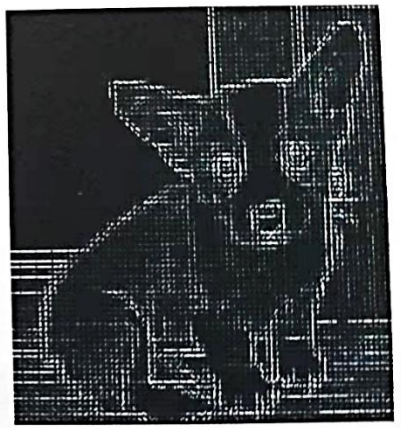


图 3-23　方向梯度直方图

2. 特徵提取

- 特徵，顧名思義是某種事物不同於其他事物的特徵。計算機爲了抓取狗的特徵，首先將相應的圖片轉化成黑白漸進的灰度圖，再把整張圖片分隔成小格。接下來統計每一個方格圖像類線條方向的分布規律，也就是沿垂直或者水平方向計算出像素變化的程度。
- 然後使用一定的算法將垂直和水平方向的梯度規程，只保留必要信息的圖像，所以終將原始圖像處理成如圖所示的樣子。
- 明顯的看出這隻狗的輪廓，並且保留了狗較長的吻部。較大且直立的耳朵等特徵信息。也需要與圖像的位置對齊，比如狗嘴和耳朵出現在圖中適當的位置。才可以對狗進行較爲準確的識別。
- 以上的圖像特徵提取方法叫做方向梯度直方圖(Histogram of Oriented Gradient, HOG)。圖像特徵提取的方法還有很多，比如基於顏色特徵的特徵識別方法。基於幾何特徵的特徵識別方法等

圖像識別

紅色符號代表狗
藍色符號代表貓

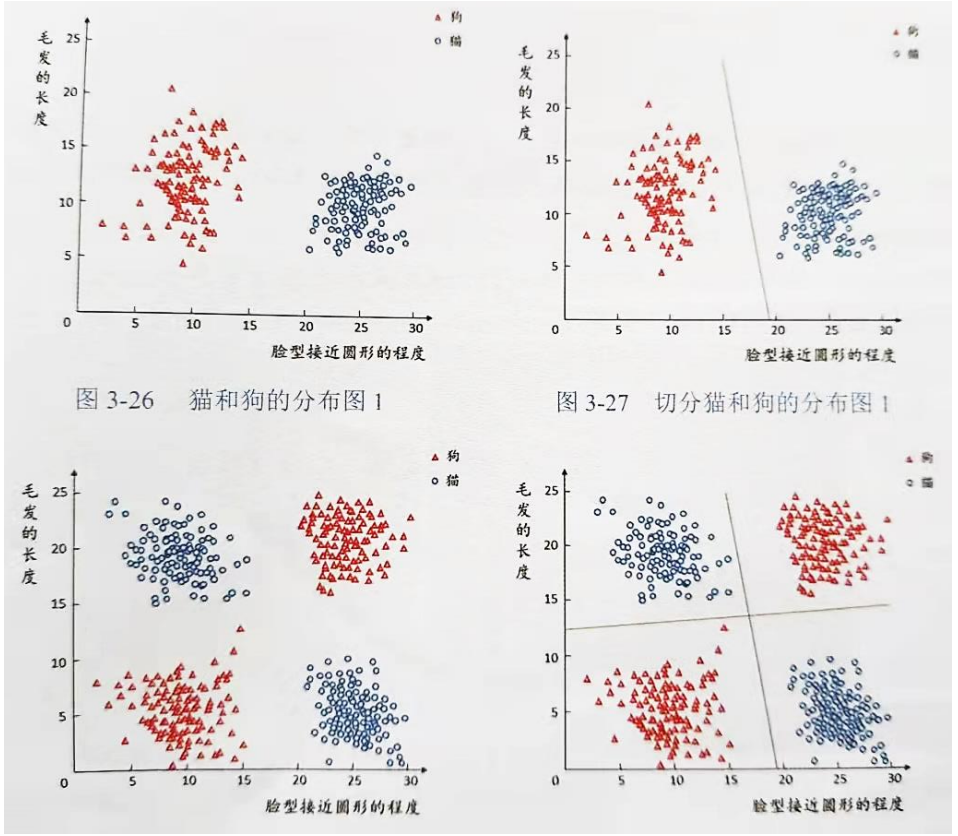


图 3-26　猫和狗的分布图 1

图 3-27　切分猫和狗的分布图 1

3. 圖像分類

- 爲了讓計算機理解圖像的內容，我們必須應用圖像分類的技術。這是使用計算機視覺以及機器學習算法從圖像中抽取意義的任務。這個操作可以簡單的理解成爲一張圖像，分配一個標籤，如貓、狗或是大象，甚至可以解釋圖像的內容並且返回一個人類可讀的句子。
- 圖像分類的依據是提取的圖像特徵，接下來我們使用圖像分類的算法對圖像進行分析。舉個例子，我們準備了一大堆貓和狗的圖片，然後用這些圖片訓練一個分類型模型。

- 如果我們要把代表狗的紅色符號與代表貓的藍色符號分開，只要在圖(3-27)中間加條直線就可以了。兩邊分屬不同的兩類，這種分類器其實就是第二章中學習過的單層感知機。如果貓和狗的分布圖有四個種類，又怎麼辦呢？這時可以使用多層感知機，也就是多層人工神經網絡。其中底層神經元的輸出是高層神經元的輸入。也就是形成了人工神經網絡中的隱藏層。

目標檢測

- 計算機面對一張圖片的時候，最基礎的任務是識別出這張圖片是什麼？其實這是一種圖像的分類。可以簡單理解成爲不同的圖片打上對應的標籤。
- 當圖片中包含一群狗的時候， 雖然計算機能夠識別出圖像中包含狗，卻並不知道圖中出現了幾只狗。這個時候就需要對圖像做進一步的處理，即進行目標檢測。

图 3-32　狗和娃娃的图片

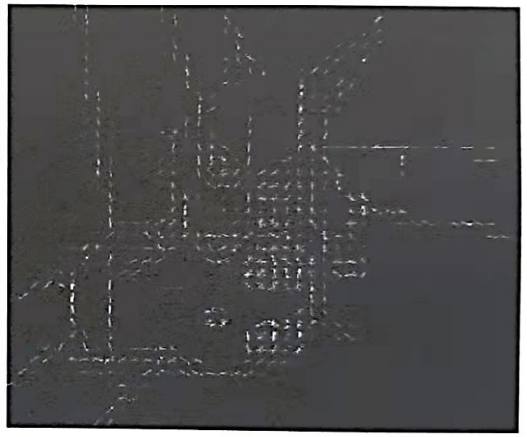


图 3-33　狗和娃娃的方向梯度直方图

圖像識別

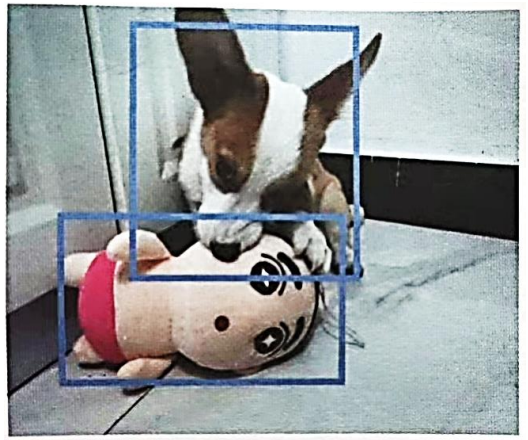


图 3-34　目标检测

目標檢測

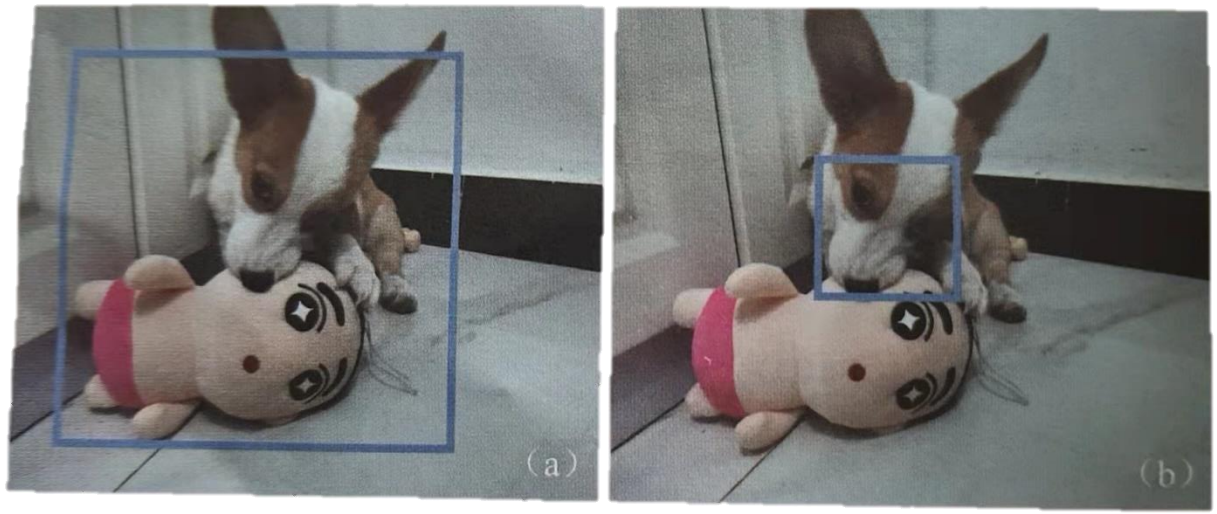
- 圖像識別與目標識別的區別: 在圖像識別中整幅圖像被分類爲單一類別，給予單一標籤。在目標檢測中，計算機需要找出圖像中目標的位置和個數，並辨認出它是什麼

目標檢測

傳統的目標檢測方法1: 傳統的目標檢測

- 傳統的目標檢測方法使用人工設計的特徵算子提取圖像特徵，用於目標檢測。我們以一種單一尺度的目標檢測方法爲例，使用固定大小的檢測框，在被檢測圖像上從左到右，從上到下移動檢測。在每一個範圍位置上檢測時，都會把這個區域的圖像輸入分類檢測算法中。然後分類選擇檢測算法會判斷這個區域內是否有物體，如果有的話，會對物體進行分類處理，一旦物體達到或超過某一分類標籤的值，就將這個檢測框標示出來，並貼上分好類的標籤 。

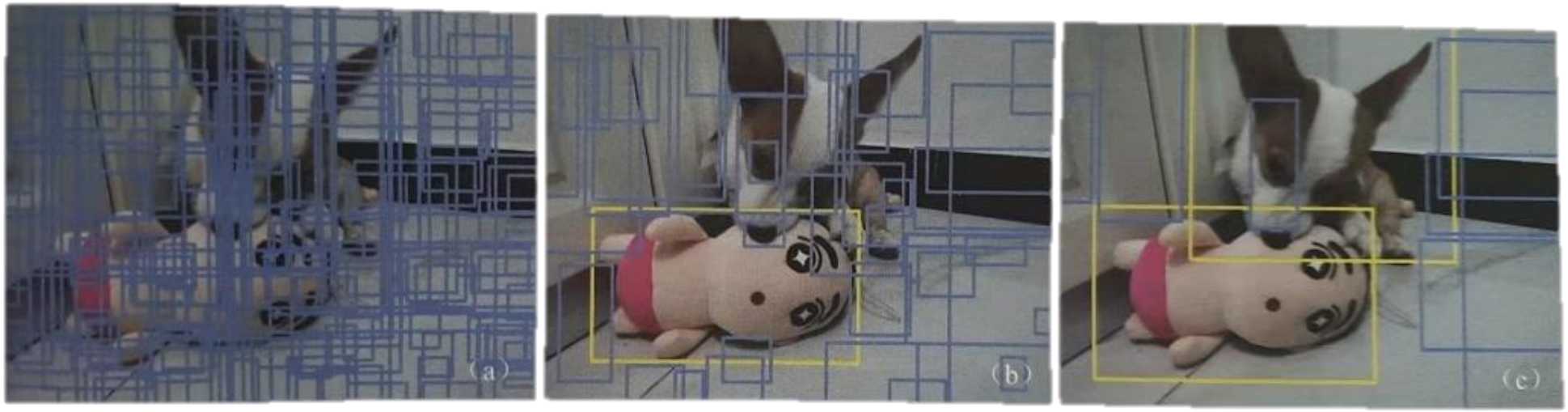
- 這種檢測也有着明顯不足，檢測框太大了，無法對檢測區域內的圖像進行分類，太小了，又無法檢測到比較大的物體



目標檢測

傳統的目標檢測方法2: 基於深度學習的目標檢測

- 基於深度學習的目標檢測。基於深度學習目標檢測方法，使用從神經網絡中學到的圖像特徵來進行目標檢測。首先，通過某些特徵框的選擇，搜索算法從待檢測圖像中提取出長上千個大大小小的目標候選框，然後把這些的候選區內的圖像分別送到神經網絡中，最後針對提取到的圖像特徵進，行分類並得到了類別信息，以及對應原圖像的位置坐標訊息，從而實現目標檢測並得到檢測結果。由此可發現，目標檢測作爲一種較爲精準的目標，其達成的難度是遠大於圖像分類的。



認識人臉識別

- “刷人臉”這個看似簡單的事情，實際上是由計算機視覺技術作爲支撐的。計算機究竟如何完成人臉採集和預處理，進行人臉偵測、人臉對齊與人臉特徵提取，實現人臉識別與或活體檢測的呢？
- 人臉採集和圖像預處理

  人臉識別過程中需要現採集人臉圖像。根據採集人臉的場景不同，一般可以分爲兩類：

  (1) 限制條件下的人臉拍攝

  (2) 非限制條件下的人臉拍攝

人臉採集和圖像預處理

- 第一類場景叫做限制條件下的人臉拍攝。比如要求坐姿標準。正對平視鏡頭，不帶帽子眼鏡，甚至還有對光線明亮的要求，拍攝證件照時，就是這一類場景，國際上的領先識別率超過99%。
- 第二類場景叫做非限制條件下的人臉拍攝。在日常生活中，自然條件下的拍到的人臉存在各種問題，比如側對鏡頭、暗光、強光、背光、遮擋、模糊等，這樣獲取到的圖像必然影響人臉算法的識別準確率。如果將第一類的場景的算法照辦到這裏，只能勉強達到60%-70%的識別準確率。所以，爲了提升算法的識別準確率，有必要從採集端着手，提升圖像採集的質量，同時可以對採集的人臉圖像進行預處理，提升圖像質量。
- 圖像增強。是一種經常採用的圖像預處理方式，例如環境光線太弱，會造成圖像變黑， 曝光不足 ，這時可以採用**暗淡部分增強、去除暗背景下的噪點、提升圖像的分辨率、去除運動時的圖像模糊**等方法進行圖像增強 。

图 3-49　曝光不足　　图 3-50　曝光过度　　图 3-51　使用 HDR 技术拍摄的照片

人臉檢測

- 人臉檢測是在圖像中準確標示出所有人臉的位置和大小。通常用一個個檢測。框標示出來。目標檢測中的“目標”其實是可以包含人臉這個目標的。
- 人臉檢測算法要解決以下幾個問題。

  (1) 人臉可能出現在圖像中的任何一個位置。

  (2) 人臉可能部分被遮擋。

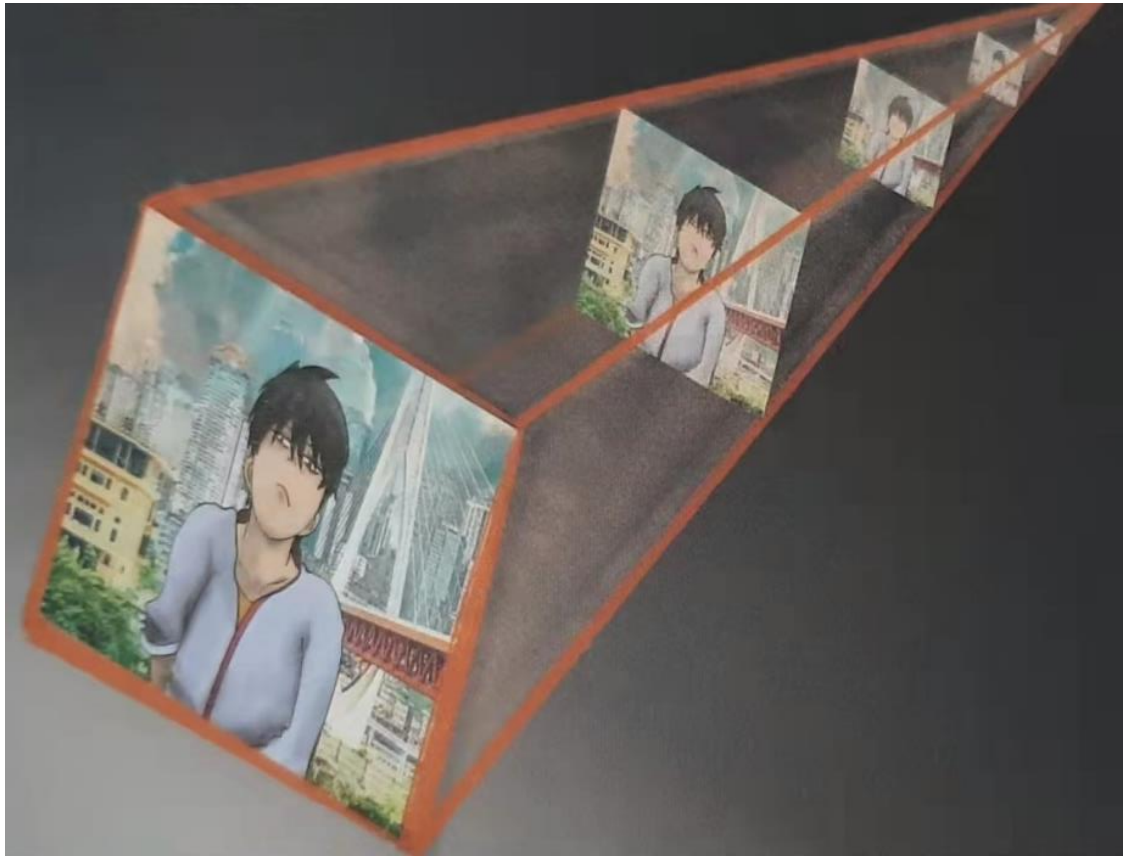
  (3) 人臉由於遠近關系可能有不同的大小。

  (4) 人臉在圖像中可能有不同的視角和姿態。

- 所以需要引入”圖像金字塔”的概念。所謂”圖像金字塔”，就是將分辨率最高的圖像放在底部，往上是一系列像素逐漸降低了圖像，外觀很像金字塔的形狀，這就構成了計算機視覺中的圖像金字塔。

人臉檢測

- 人臉檢測過程中，人連圖像金字塔顯得尤其有用。因爲人臉在圖像中會有遠近，大小之分。人臉檢測算法會規定一個最小的檢測範圍，比如12x12的像素，小於這個範圍的人臉就不偵測，大於這個範圍怎麼辦？就把檢測到的圖像逐步降低分辨率，直到12x12像素在這個過程中，如果有檢測到人臉，就標記上檢測框。計算機會把表示出來的檢測框進行合並去重，保留最合適人臉尺寸的檢測框。



人臉特徵提取

- 計算機檢測到人臉之後，根據輸入的人臉頭像，尋找並定位出面部關鍵的特徵點，比如眼睛、鼻尖、嘴角點、眉毛以及人臉各部分輪廓點等。
- 就是所示的綠色的點，如果人臉圖像因爲角度、姿態關系有些歪的話，可以根據關鍵特徵點，使用變換的方法把人臉"對齊"或者說進行"擺正"，提高識別準確率。
- 使用HOG特徵算法進行圖像的特徵提取.

3-53 人脸关键特征点　　图 3-54 HOG 特征算法

綠色的點

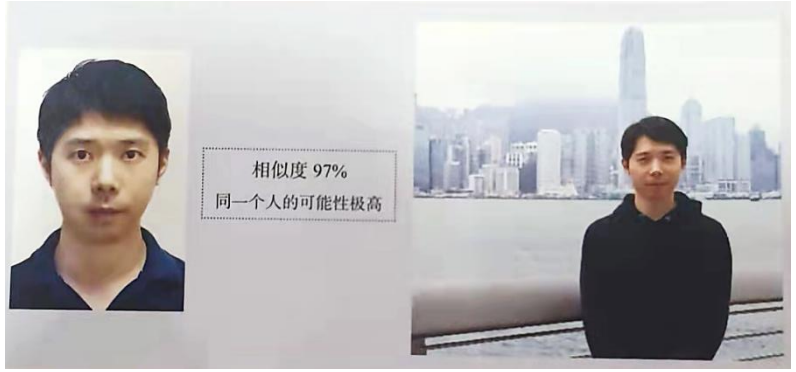
人臉對比

- 我們通過前面的步驟獲取的人臉特徵。目的就是與人臉數據庫中的圖像信息進行對比。換句話說，人臉識別就是將待識別的人臉特徵與已得到人臉特徵進行比較，根據相似度、對人人臉身份信息進行判斷。 可以把”比較”分成兩類: 一類是一對一進行圖像比較的過程，可以叫做確認，通俗說就是證明兩張人臉圖像是否屬於同一人。另一類是一對多進行圖像匹配對比的過程，可以叫做辨認，通俗說是。能不能找到這個人。

(1) 第一類比較場景中，提取人臉的特徵之後。與事先準備好的人臉特徵進行對比。只要兩者對比的相似度，在一個確定的值範圍內，我們就確認了兩者是同一個人。

(2) 第二類比較場景中，將提取到的需要比對的人臉特徵值，與人臉數據庫中所有人臉的特徵值進行對比。也是通過攝像頭從實時採集到的人臉圖像中提取的。這對人臉識別預處理要求較高 。

→ 相似度與值範圍

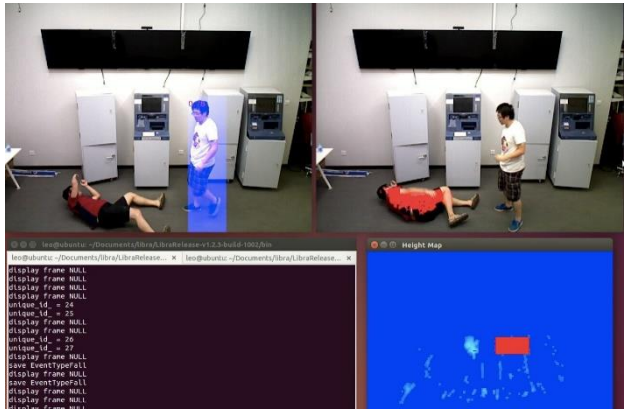


→ 跨年齡人臉識別

計算機視覺技術更大範圍的應用

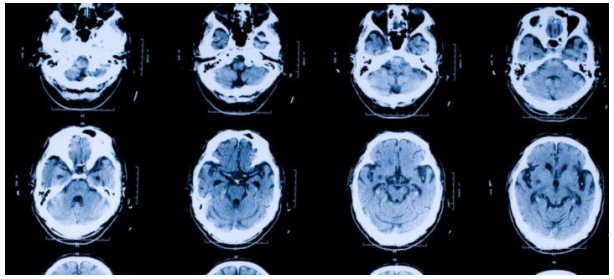
- 計算機視覺與工業製造

- 計算機視覺與安防監控



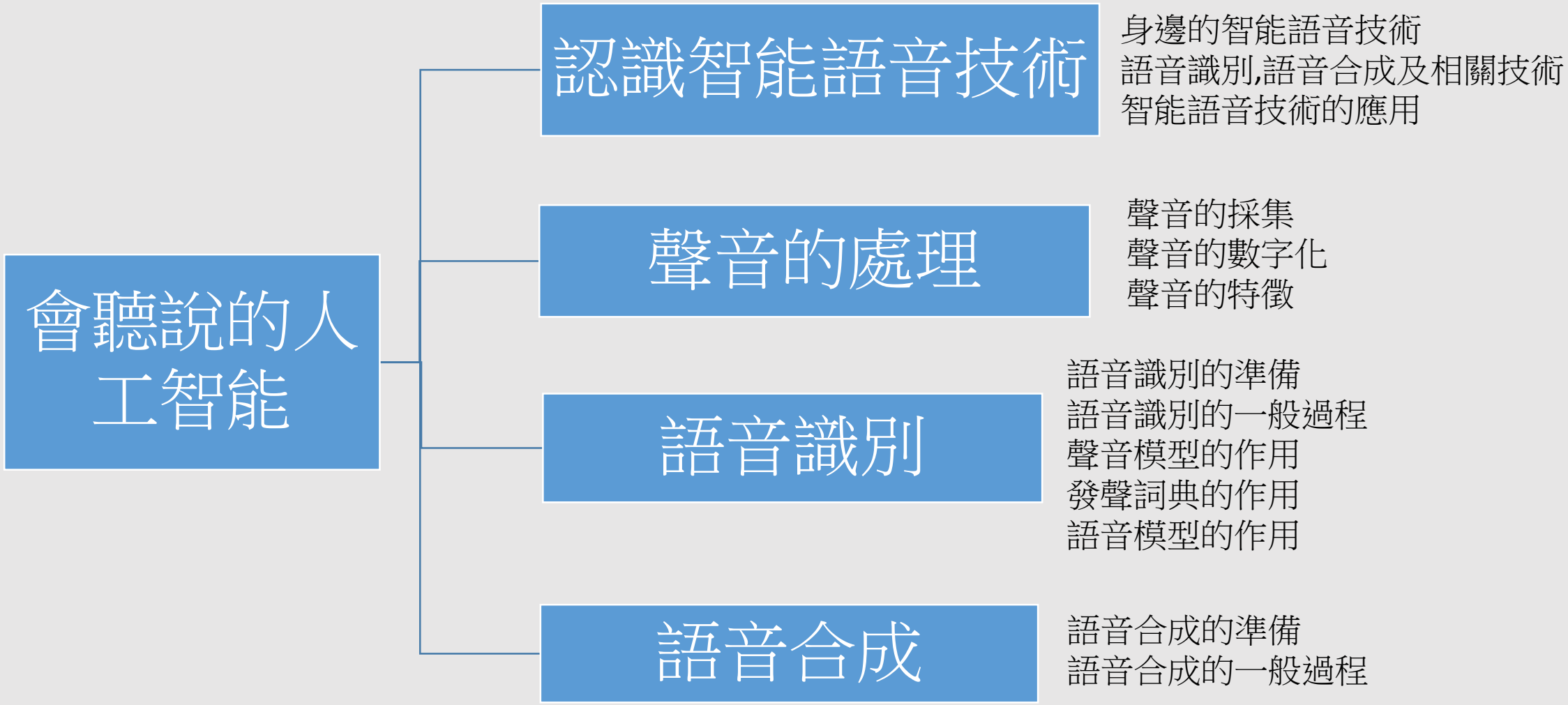
- 計算機視覺與農業

- 計算機視覺與醫療



- 計算機視覺與自動駕駛

(第四章) 會聽說的人工智能



Edit by Hammond Lai

身邊的智能語音技術

- 大致說來，手機通過語音識別技術聽見了裏的問題，結合自然語音處理技術和相關應用程序，理解了你的問題，並找到了問題答案，在通過語音合成技術把答案說給你聽，這裏的語音技術識別、語音合成技術都屬於智能語音技術。

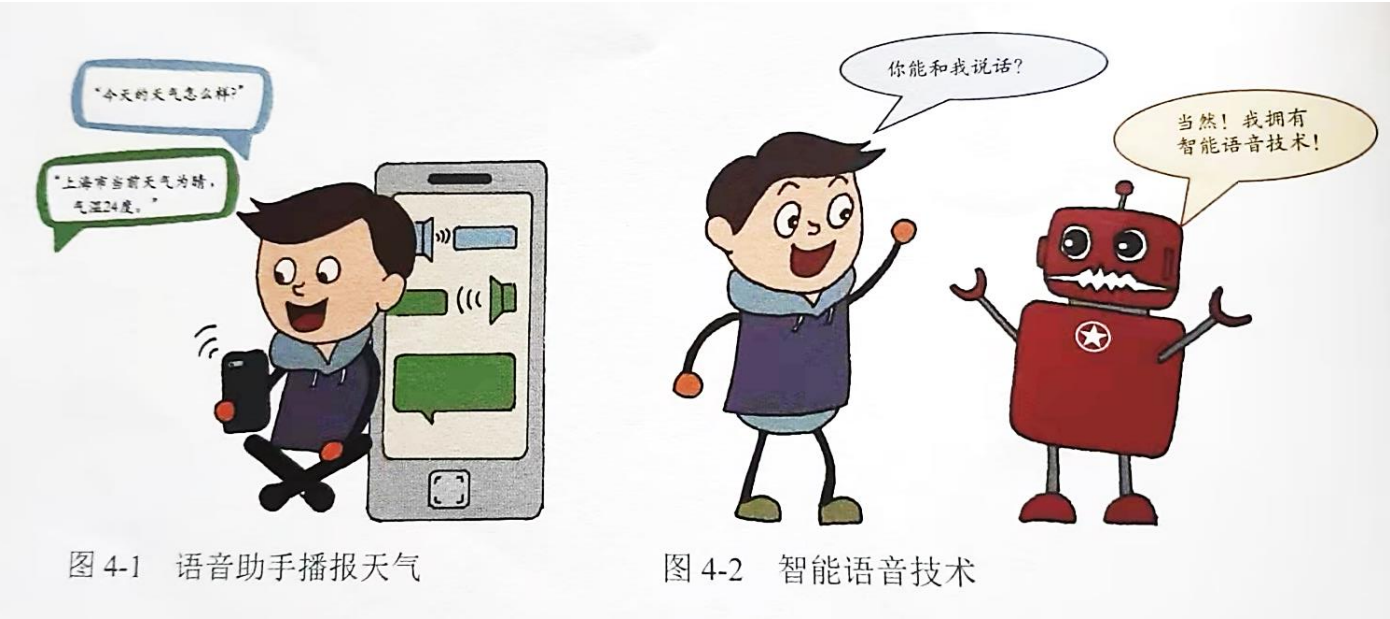


图 4-1　语音助手播报天气

图 4-2　智能语音技术

語音識別,語音合成及相關技術

- 語音識別技術是一種可以讓機器”聽得見”的技術，可以把我們說的話轉變成文字，方便機器去理解。不過在開始語音識別前，需要先對聲音進行採集和數字化處理。這些我們在下一節具體介紹。
語音合成技術是一種可以讓機器”說的出”的技術，可以把文字轉變成語音。

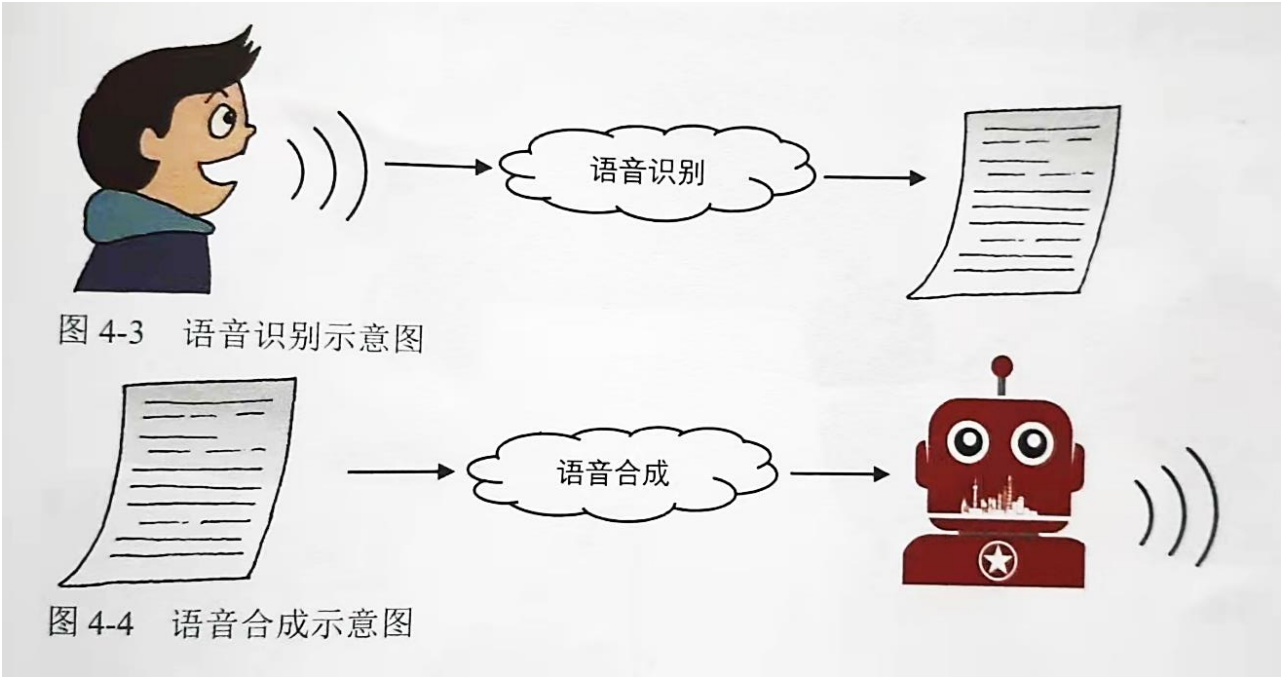


图 4-3　语音识别示意图

图 4-4　语音合成示意图

語音識別,語音合成及相關技術

- “聽得見”不代表聽得懂，”說得出”也不代表知道說什麼。所以在使用語音識別技術獲取對話中的問題文字後，還需要運用自然語音處理技術對問題文字加以理解，並生成應答文字。在使用語音合成技術把應答文字說出來。才能真正實現我們的對話，做到既”聽得見”又”聽得懂”，既”說得出”，又”說得對”。

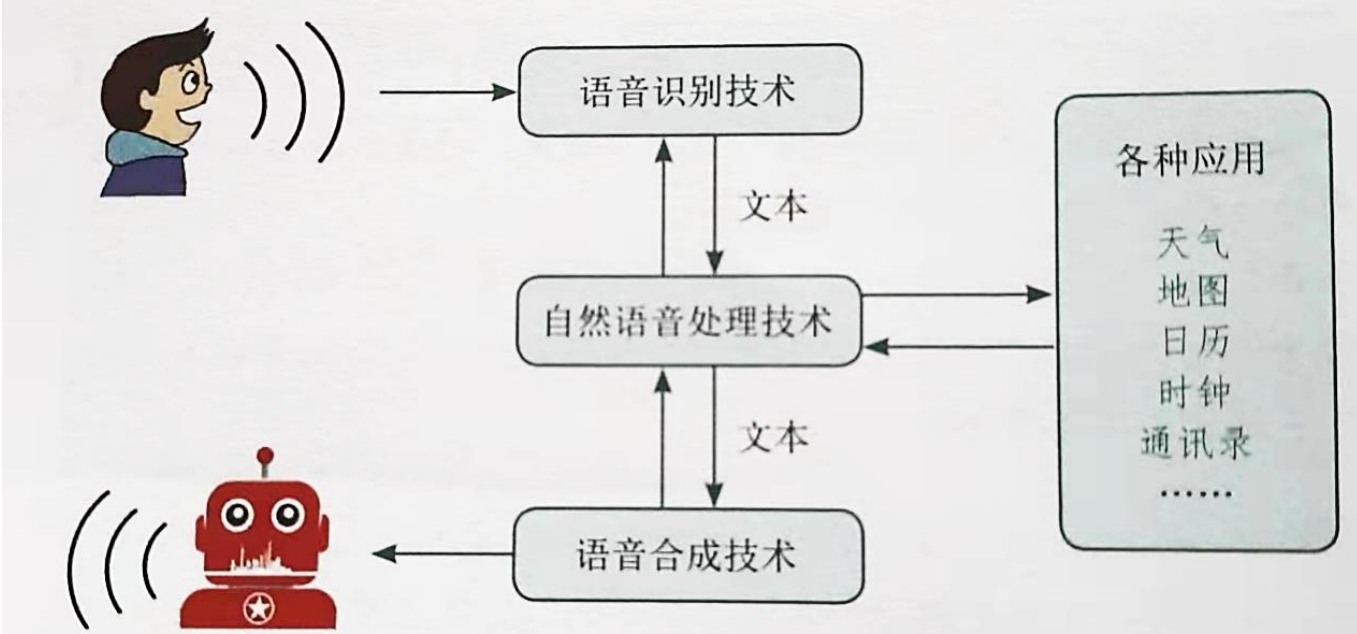
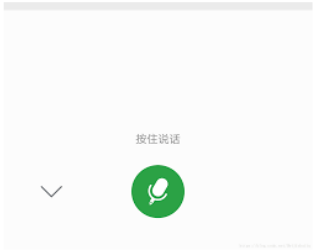


图 4-5　人机语音对话示意图

智能語音技術的應用

- 智能語音技術在我們的日常生活中得到了廣泛的應用。語音助手就出現在。智能手機、智能音箱、平板電腦等許多地方。我們可以用說話的方式查詢天氣、設置鬧鍾，甚至與他聊天逗趣。

1. 語音翻譯
   語音翻譯及口譯。把一種語言的口語翻譯成另一種語言的口語。在許多國際會議上，經常需要進行多種語言之間的口語即時翻譯。
2. 語音輸入
   語音輸入及用說話的方式來輸入文字。相比傳統的拼音、筆畫等輸入法，可以減少對鍵盤的使用，提高輸入速度，更加簡單和方便。
3. 語音控制
   語音控制及用語音發出命令，控制各種設置作出反應。比如智能家居中用到的語音開關電視。調整空調溫度等。
4. 語音客服
   語音客服接替人工客服完成接聽電話的工作，協助顧客完成業務查詢、業務辦理等需求，減少人工客戶客服的工作量，並可以24小時不間斷地提供服務。

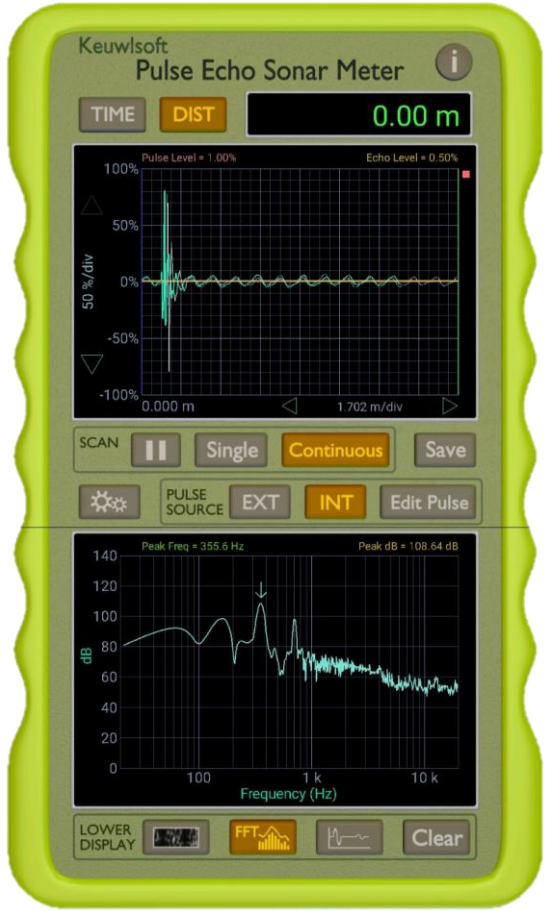


聲音的採集

聲音是通過物體振動產生的，需要通過介質才能傳播，而空氣就是最常見的一種介質。那麼聲音在傳播的過程中，怎麼把它採集起來並進行保存呢？首先需要用到的是麥克風，也就是我們平時說話的話筒。不同類型麥克風的工作原理是不一樣的，但基本的用途都是將傳入的聲音轉化成電信號，只是轉化的個方式各不相同。

各種麥克風

話筒收集聲音



聲音轉化成電信號

聲音的數字化

- 聲音經過麥克風轉化成電信號以後，再可以經過簡單處理，用模擬信號的方式保存起來， 進一步轉化爲數字信號，供計算機處理使用，這個過程就稱爲聲音的數字化。在計算機中處理的信息都是採用了二進制的 ”0”和”1” 來表示，聲音信號也不例外。聲音信號的數字化過程需要經歷採樣和量化。

1. 聲音採集

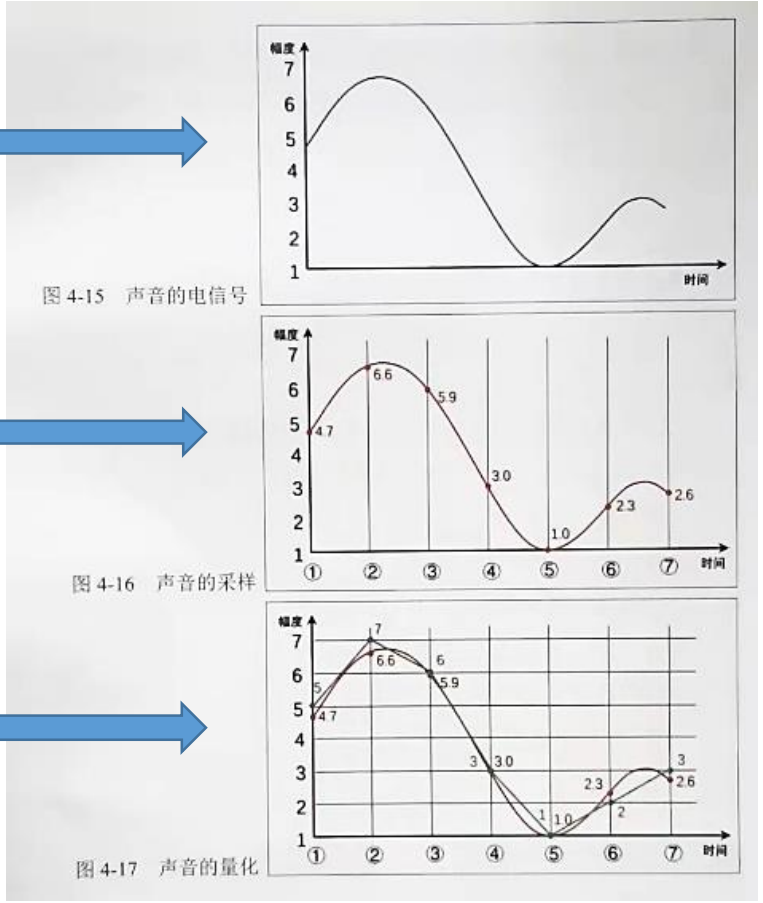
可用麥克風接受聲音轉換成電信號

2. 採樣

經過麥克風採樣的聲音。轉化成的電信號是一種連續變化的模擬信號。如果用圖的方式來表示的話，可以把它想象成一條曲線。這條曲線可以理解成是由許許多多個點組成的，每個點的信息都記錄着當前時間、聲音、振動的幅度。

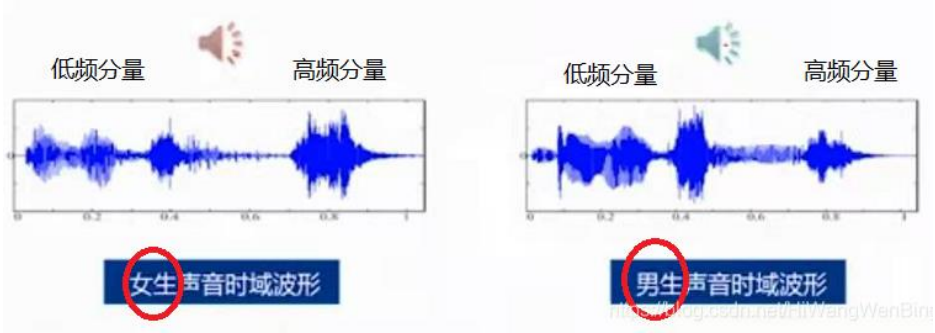
3. 量化

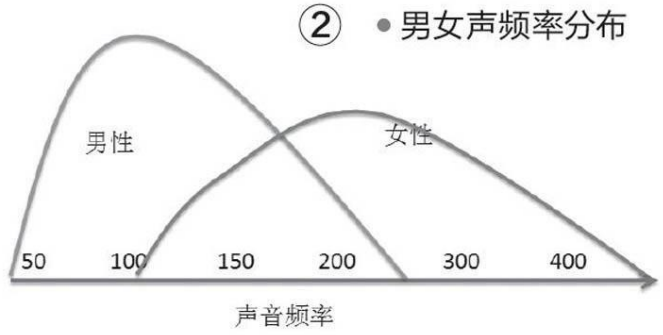
經過採樣，得到了若干個採樣點的信息，每個點記錄的聲音幅度都來自模擬的電信號，是一個比較精確的值。這些值各不相同，給之後的處理帶來了困難可以用類似四舍五入的辦法得到新的一組值，這個過程叫做聲音的量化 。



聲音的特徵

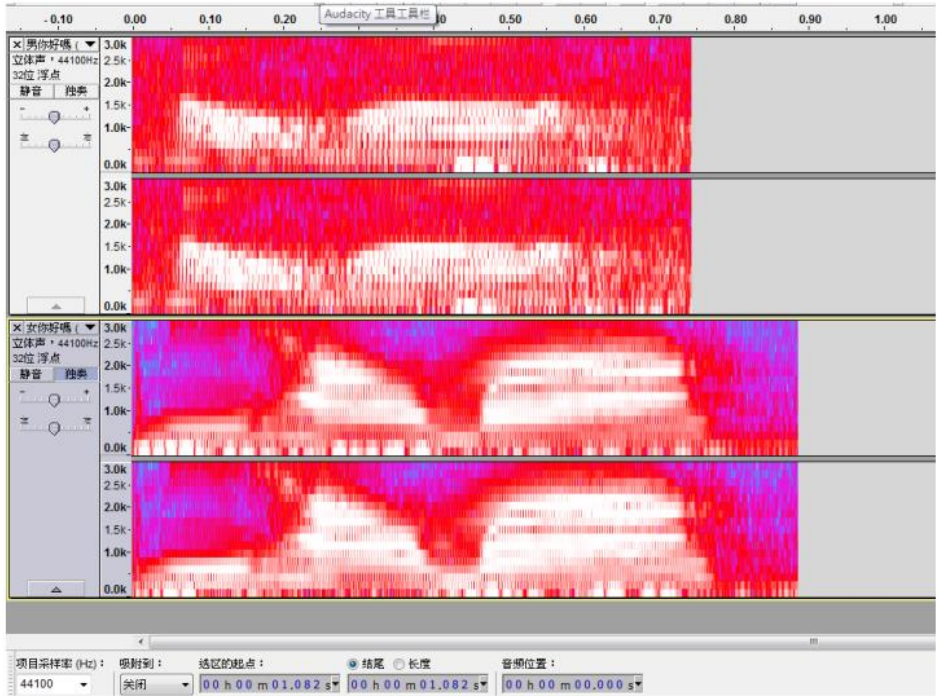
- 計算機可直接處理的”0”和”1”這樣的數字信號後，爲了方便人們在計算機上對聲音數字信號進行觀察和編輯，大多用波形和頻譜的方式來表示。不同人說同樣的話，發同樣的音。聲音的波形是不一樣的。這是因爲不同人說話發出聲音，有着不同的特徵。
- 各人有著不同的聲音特徵: 音量不同，聲調不同，音色不同。





男聲

女聲



語音識別的準備

- 語音識別技術可以把我們說的話轉變成文字。究竟是怎麼實現的呢？其中經典的語音識別算法是有聲學模型、語言模型和發聲詞典組成的。我們就以他爲例，來看看說話的語音是怎樣一步步變成文本文字的。

- 首先，用大量聲音數據訓練一個聲學模型，以用來將聲音轉換爲聲學符號。

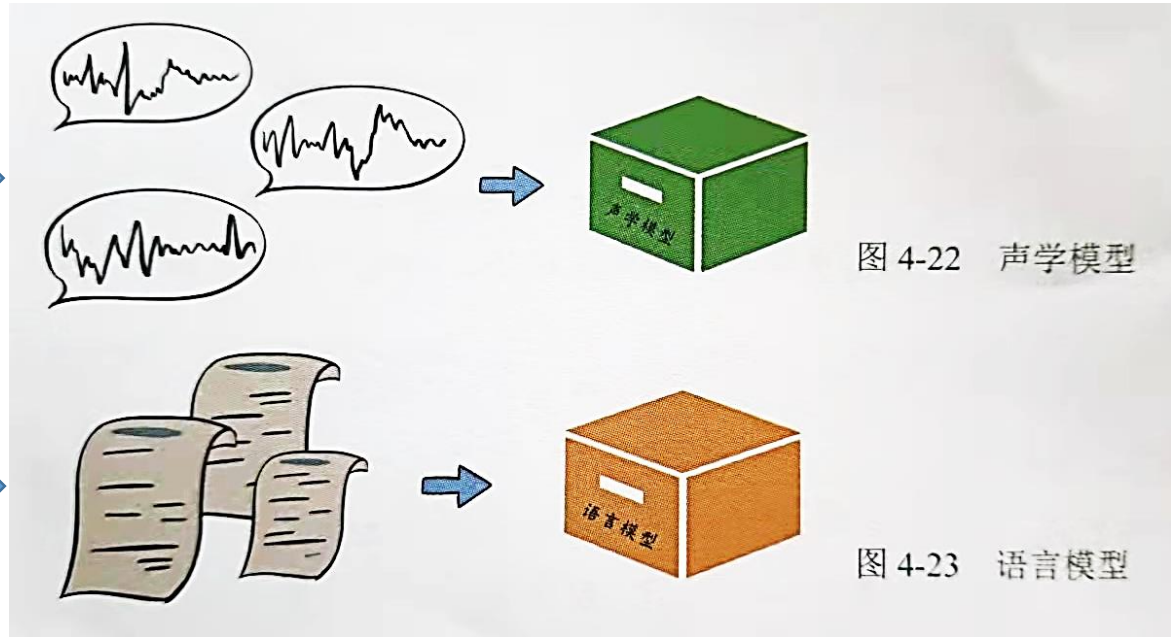


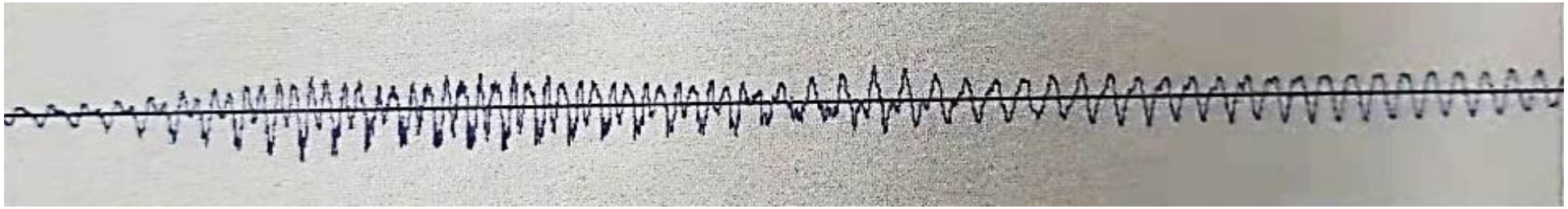
图 4-22　声学模型

- 接着用大量文本數據訓練一個語言模型，爲聲學符號找到可能的文字表達

图 4-23　语言模型

語音識別的一般過程

- 比如把人工智能這四個字讀音的波形圖放大如圖，由於太長，這裏只截取顯示了一部分。



- 開始語音識別前，需要將波形進行一些預處理。比如錄音時的環境噪音，沒說話時的空白、靜音等多餘的信息，預處理就是要把這些信息剔除。方便接下來的特徵提取操作。得到預處理好的聲音後。就可以開始聲學特徵提取了，如提取梅爾頻率倒譜系數(MFCC)。感知線性預測(PLP)。

- 語言由單詞組成。單詞由音素組成，因此對音素的識別可以說是語音識別技術的關鍵。

語音識別的一般過程

- 了解了什麼是音素。接着我們來看語音識別的一般過程。
- 第一步。對聲音進行分幀。也就是把聲音切成一小段一小段，每小段稱爲一幀，分幀的操作並不是簡單的切開。而是按照一定的時間間隔，切割成彼此送重疊的幀。

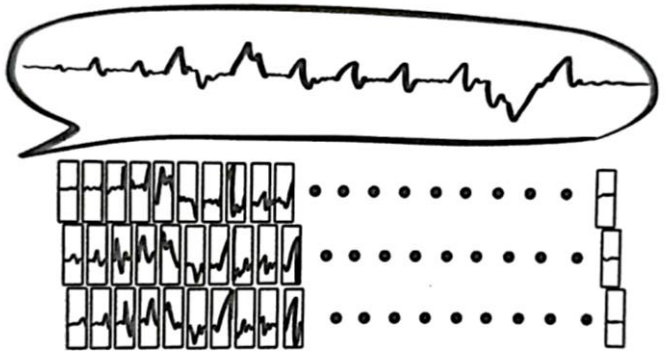


图 4-25　分帧的过程

- 第二步，通過聲學模型。將這些幀爭識別成對應的語音狀態，也就是用幀組成狀態。這裏所說的狀態可以理解成比音素更細致的語音單位。到底多少幀對應一個語音狀態呢？由於每個人的音量、音調、音色都不同，因此，把幀識別成狀態是語音識別技術中的一個難點。通過訓練好的聲學模型去判定，看哪些章對應某個狀態的概率最大，那麼這些幀就屬於某個狀態。
- 第三部。用狀態組成音素。一般用三個。狀態組合成一個音素。

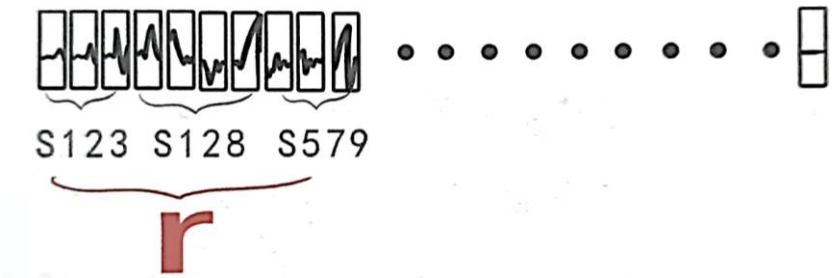


图 4-26　把帧识别成状态，用状态组成音素

語音識別的一般過程

- 第四步。把音素組成組合成單詞。在這個過程中，聲學模型仍然非常重要，它能找到音素最可能的聲學符號表達。結合發聲詞典，將聲學符號與語言模型中最有可能的文字表達相對應。就能將聲音轉換成文字。

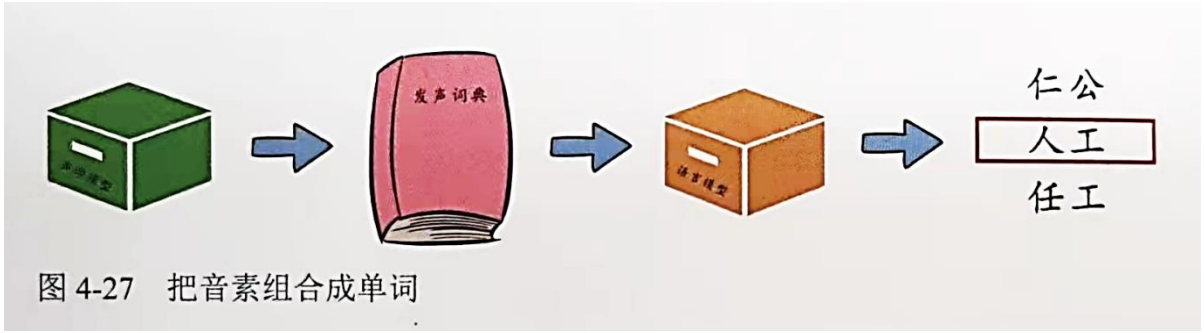
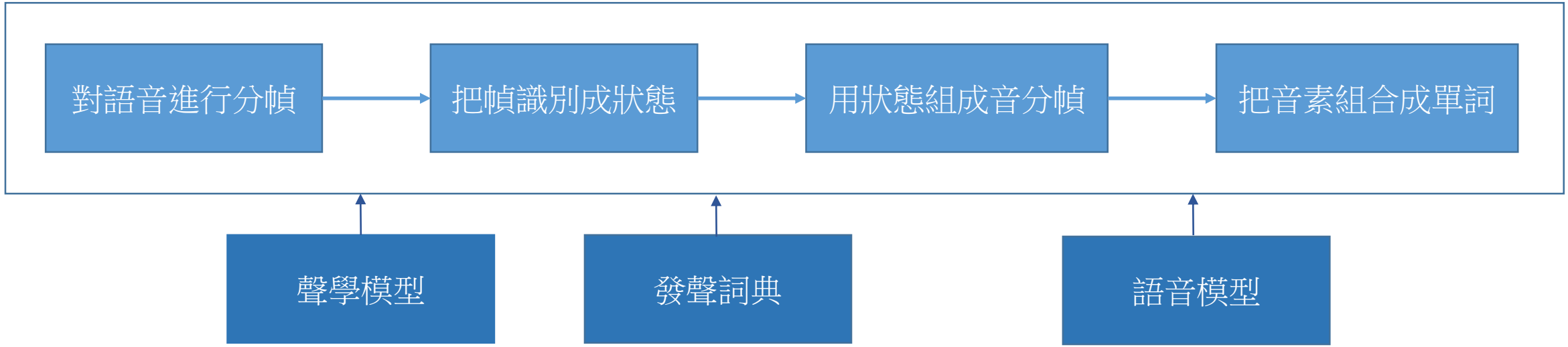


图 4-27　把音素组合成单词

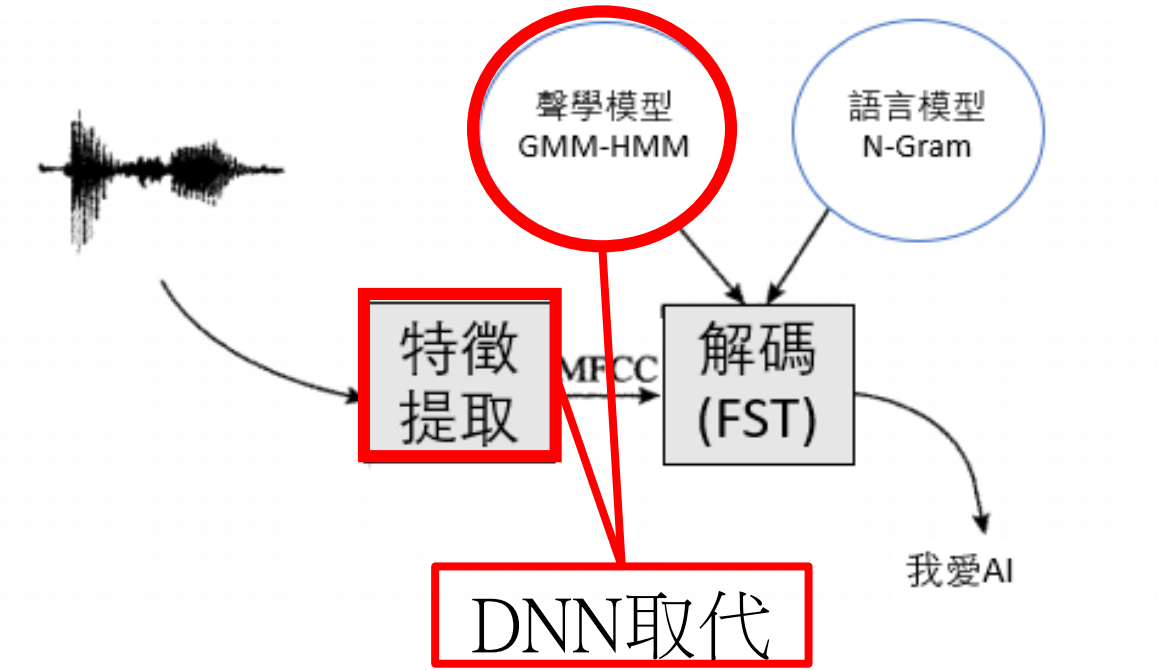


Edit by Hammond Lai

聲學模型作用

- 聲學模型的主要作用就是把語音分成的幀識別成對應的狀態，把狀態組成音素，將音素對應到最有可能的聲學符號表達。是可以簡單理解爲通過大量的語音數據來訓練聲學模型，從而更好地完成這些任務。

- 訓練聲學模型的方法也有很多。傳統的方法主要是採用混合高斯模型的隱馬爾科夫模型(GMM-HMM)，現在的方法則是使用深度神經網絡的隱馬爾科夫模型(DNN-HMM)



發聲詞典的作用

- 發聲詞典也叫發音字典，不同的語音有着不同的發聲詞典，例如漢語普通話的發聲詞典，裏面寫的是漢語拼音與漢字的對應情況。而英語的發聲詞典裏就是音標與單詞的對應情況。
- 還是以漢語普通話爲例，發聲詞典的主要作用是根據聲學模型識別出來的音素，找到對應的漢語拼音和相應的漢字，在聲學模型和語言模型中間建立連接，把兩者聯系起來。

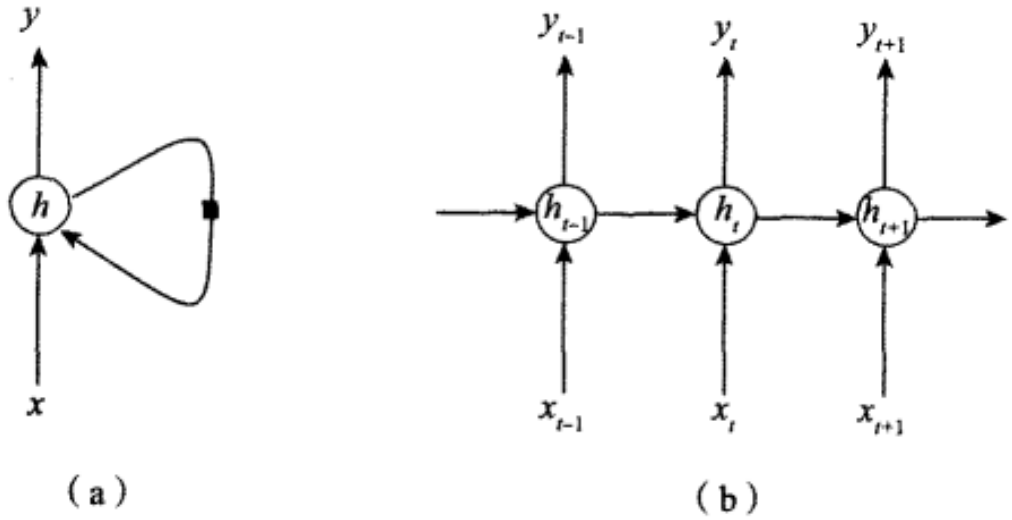
字詞	發音
一	yi1
一下	yi2 xia4
一下子	yi2 xia4 zi

語言模型的作用

- 漢語相同的讀音會有很多不同的字和詞，那怎麼知道究竟哪個字詞是正確呢？就拿“rengong”爲例，經過聲學模型和發聲詞典的匹配後，可能得到同音的。詞語由仁公、人工、任工。
- 例如在”這是一種人工打標籤的方法”的語音中。機器該選哪個詞語呢？這些詞語在我們來看，很容易挑出正確的答案，但對機器來說卻沒有什麼差別，這時就需要語言模型的幫助。它的主要作用就是從這些同音字詞中，根據人們的語音習慣，篩選出最符合語句原本意思的字詞，再將字詞組合成句子。

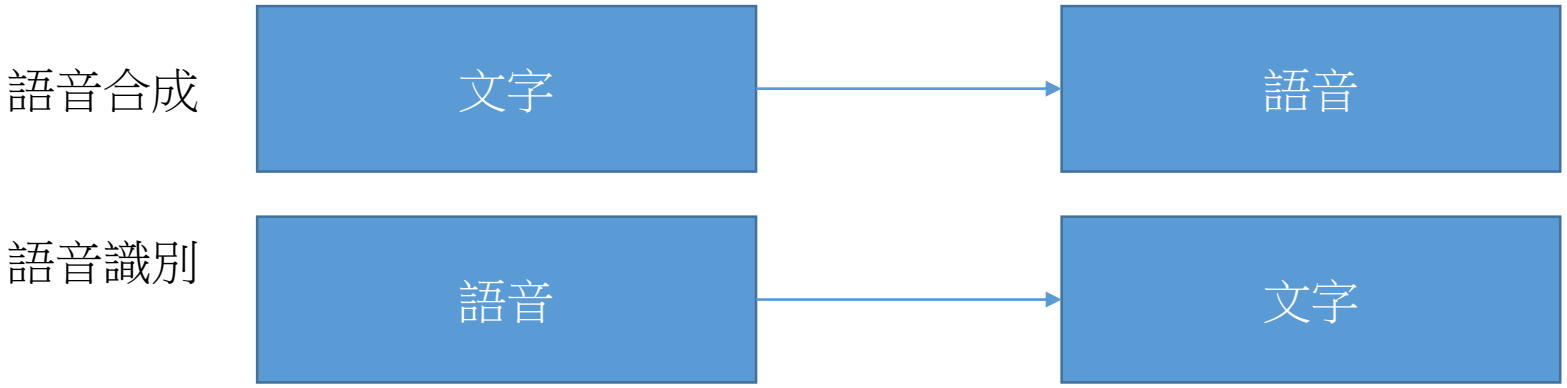
循環神經網絡

- 循環神經網絡(RNN)是神經網絡的一種，與傳統的神經網絡相比，每層的神經元之間也有連接。這樣的結構對於理序列化的數據很有優勢。對一個序列而言，當前的輸出與前面的輸出有關。而語音識別時講話的語音就是有時間序列的數據，往往前面的話和後面的話是有關聯的，也就是我們平時說話的聯系上下文，所以運用循環神經網絡對語音數據和文本數據進行學習訓練，生成聲學模型和語言模型，就可以對輸入的語言進行很好的識別。



語音合成的準備

- 語音合成技術是一種讓機器”說得出”的技術，可以把文字轉變成語音，聽上去好像和語音識別正好相反。那麼語音合成是不是只要把語音識別的過程反過來就可以呢？可能還是有些不一樣。
- 在語音識別之前需要做一些準備工作，在語音合成之前同樣需要做一些準備工作。例如錄制一些語音作爲語音庫，爲了能夠覆蓋語音中的各個音素與各種音調等，錄制的內容一般需要經過專門設計，而根據合成聲音的不同方法，語音庫的大小也不相同。從數小時到數百小時不等。



語音合成一般過程

- 語音合成的一般過程可以分爲以下兩步。
- 第一步，**預測文本的讀音**。我們仍在朗讀文本的時候，會對文本進行分析，比如聲調是怎麼樣的，怎麼樣分詞、重音在哪裏，節奏怎麼樣等。機器在語音合成時，首先要做也是分析文本，這樣才可以讓生長的語音更加自然。這裏一般需要用到自然語音處理技術，在完成文本分析等工作後，再把需要合成語音的文本信息轉換成語音素序列。
- 第二步，合成聲音。完成這一步的方法有很多。這裏簡單介紹其中兩種。一種是**波形拼接法**，顧名思義，這是一種把聲音波形拼接在一起。形成所需語音的方法。具體來說，就是根據之前第一步形成的音素信息，到語音庫中尋找與之匹配的聲音，進行必要的調整後，把這些聲音波形拼接起來，完成聲音的合成

学 习 人 工 智 能 非 常 快 乐

xuexi|rengongzhineng|feichang|kuaile

图 4-30　预测文本的读音

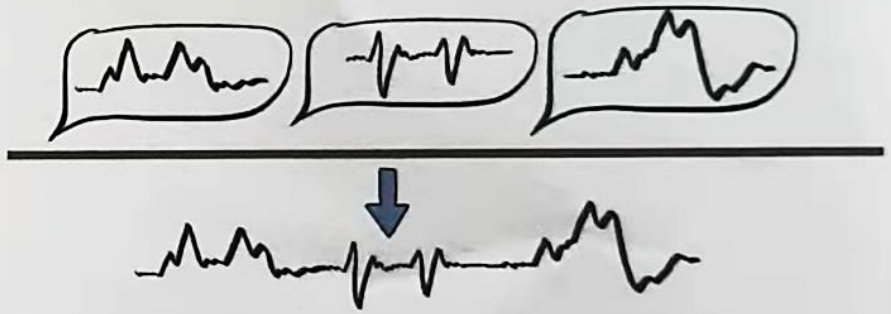


图 4-31　波形拼接法

語音合成一般過程

- 另一種方法是統計參數合成法，它會根據之前第一步形成的音素信息，先將其轉換成連續的語音參數。然後結合從語音庫中提取到的聲音特徵，運用相關算法生成相應的語音。
- 這兩種方法各有優缺點，在實際使用時往往會將兩者相結合，可以達到更好的合成效果。
- 隨着機器深度學習技術的發展。端到端的語音合成技術日趨成熟。所謂端到端，指的是只要一端輸入文本，另一端就可以直接輸出語音，在中間過程中使用深度學習技術對聲學模型進行訓練和應用 。微信就是把這語音合成技術廣泛使用.

完

Edit by Hammond Lai