計畫名稱

社交型協作機器人基於情境的涵意提供適切的服務

摘要

本文的目的是提出一種社交型協作機器人應用情境的涵意,學習並預測人類的想法,進而提供「剛剛好的服務」。為了在人類社交環境中與人友善的互動, 機器人應具有情境感知瞭解人類社交技巧的能力並且表現出得體的行為。

在本文中,情境式上下文著重在讓機器人感知他人是否需要幫助,根據預測的人類想法,機器人提供剛剛好的服務。剛剛好的概念來自鼎泰豐餐廳的董事長, 他說:「服務不足,是怠慢;殷勤過頭,變成打擾,『剛剛好的服務』是鼎泰豐團隊努力追求的目標」。在服務業方面,當顧客需要幫助時,服務員主動提供服 務是非常暖心的。換句話說,當顧客不需要幫助時,不去打擾他們是很體貼的。

我們提出兩個深度學習模型,作為機器人的情境式上下文感知,並從人機互動中觀察並學習判斷人類的意圖。基於深度學習模型,我們賦予機器人感知人的意向的能力。因此,機器人可以基於預測的人類心理狀態,做出適當的社交行為。實驗結果表明,與常規分類器相比,我們提出的深度學習模型可以使機器人顯著 提高預測人類思維的準確性。此外,在判斷人是否需要幫忙的任務上,基於情境 式上下文的預測結果與服務業人士的意見保持高度一致。

研究的問題所在及研究動機

本文的目的是開發一種社交型協作機器人，以提供“剛剛好的服務”，使用基於情境環境的感知來感知人的思想。在不久的將來，越來越多的人性化機器人將參與到人類的生活中。為了在人類社會環境中更加友好，機器人應該具有社會語境知覺的能力。在本文中，我們提出兩個深入學習模型，從以前的人機交互觀察中學習，以使我們的機器人能夠感知人的需要。因此，對人類狀態的適當社會行為將由機器人作出反應。實驗結果表明，提出的深度學習模型可以顯著提高從人類相互作用行為預測人的參與的準確性。此外，我們的社交型協作機器人的預測與服務行業人士的意見高度一致。

技術研究方法及創新性

特徵擷取

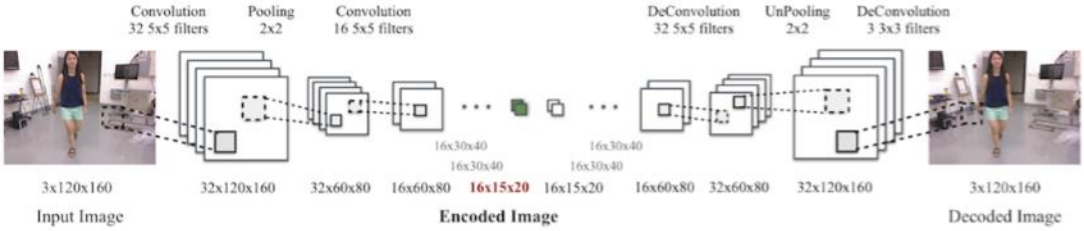

圖 1 卷積式auto-encoder 基於卷積式類神經網路.

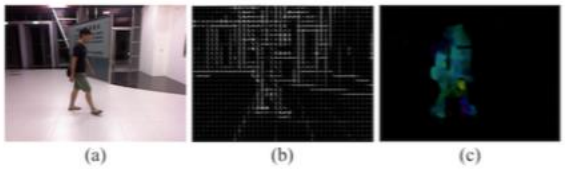

圖 2 展示手工特徵。(a) 是原始的原始圖像。(b) 顯示了定向梯度直方圖 (HOG) 特徵。 (c) 顯示光流特徵.

圖1顯示我們使用Deep Learning 的一個分支CNN(Convolutional Neural Networks) 做了一個auto-encoder 目的是希望可以幫助我們抽取人類姿態的資訊並降取資料維度，圖2則是我們選擇的一些分係人類姿態的特徵來與我們的auto-encoder 做比較。

演算法架構

循環神經網絡(RNNs)是序列學習中最著名的模型。 與傳統的前饋神經網絡不同，RNN 具有循環連接，使得它們能夠從先前輸入的全部歷史映射到目標向量。

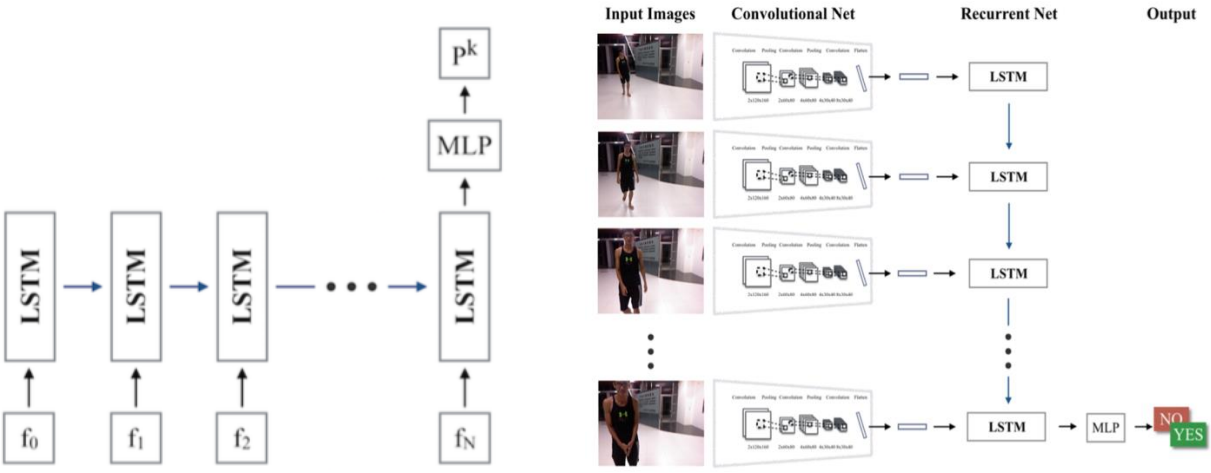

圖 3 (a) LSTM-based 模型 (b) CNN followed by LSTM 模型

在監督學習中，RNN 可以通過反向傳播時間（BPTT）與順序輸入數據和輸出目標進行訓練。 然而，模型訓練 BPTT 期間梯度爆炸或消失的問題阻礙了 RNN 的表現。這意味著基本的 RNN 可能不會處理長期依賴[23]。因此，長時間內存網絡（LSTM）是一種建議，以防止這些問題。 一個 LSTM 單元由三個門和單元狀態組成，LSTM 單元的方程如式（1）-（5）所示 (6)則是我們定義的 cost function。

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf} + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (2)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (4)$$

$$h_t = o_t \tanh(c_t) \quad (5)$$

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{n=1}^{N}[y^n \log\hat{y}^n + (1 - y^n)\log(1 - \hat{y}^n)] \quad (6)$$

model (b) 與 model (a) 最大的差別在於在model (a) 我們需要分別學習空間和時間因素，這意味著我們提取空間特徵，然後應用LSTM來順序學習。 在最後一個實驗中，我們想在一個模型中，從深入學習的模型自動學習時空特徵。 這意味著CNN隨後是RNNs，增強了學習架構，如圖1所示。 7可能潛在地保持提高準確性的能力。 因此，增強的學習架構以原始圖像，HOG圖像和光學流量圖像作為輸入。

研究成果

演算法評估（五倍交叉驗證）

	LSTM-RNN (Our Approach)			SVM			Gaussian Naive Bayes		
	Encoded Image	HOG	Optical Flow	Encoded Image	HOG	Optical Flow	Encoded Image	HOG	Optical Flow
Training Accuracy	86%	99.37%	97.5%	53.38%	54.25%	95.13%	98.63%	99.88%	68.25%
Testing Accuracy	**73%**	**69%**	63%	49.5%	49.5%	**66%**	64%	60%	57.5%

Table 1

	LSTM-RNN (Our Approach)			SVM			Gaussian Naive Bayes		
	Encoded Image + HOG	Encoded Image + Optical Flow	Optical Flow + HOG	Encoded Image + HOG	Encoded Image + Optical Flow	Optical Flow + HOG	Encoded Image + HOG	Encoded Image + Optical Flow	Optical Flow + HOG
Training Accuracy	94.75%	88.74%	98.12%	53.25%	91%	91.125%	96.625%	84%	74.375%
Testing Accuracy	**75%**	66.5%	64%	45.5%	**68.5%**	**67.5%**	64%	63%	60%

Table 2

	Raw Image	HOG	Optical Flow
Training Accuracy	50.5%	48.0%	92.13%
Testing Accuracy	46%	47.5%	**78%**

Table 3

來自 Table 1 的結果，應用 LSTM-RNN 架構，SVM 和高斯樸素貝葉斯從上述特徵中分類人類需要幫助。Table 2 顯示應用 LSTM-RNN 架構,SVM 和高斯樸素貝業斯將上述兩種特徵接在一起做多種特徵輸入的比較。來自 Table3 的結果通過 5 次交叉驗證，應用 CNN 接 LSTMs 架構，以從不同的輸入圖像中學習。我們發現，機器人可以通過加入時間與空間中的資訊大幅提升辨識率，進而提供剛剛好的服務

社交型協作機器人判斷 ＶＳ 服務業專業評估

圖 4

我們想分析服務行業人們對社會機器人的預測與決策之間的相似性。因此，我們使深入學習 model 2 (CNN followed by LSTM)即時的運行，從而機器人可以觀察一個人的行為，並實時預測他/她的表情。兩個觀察結果如圖 3 所示。當一個人出現時，機器人首先開始問候，開始觀察大約五秒鐘的人的行為。那麼機器人就會對人類思想的預測做出適當的反應。如果只有預測人類的思想是真的（即人們可能需要一些幫助），機器人會主動禮貌地說"我可以幫你嗎？"否則，機器人將保持沉默，以免打擾人。最終共有 29 個人機交互。然後，這些意見也由 17 名服務工作人員進行評估，其中包括鼎泰鋒員工，餐廳主廚，餐廳老闆，醫院實習生等。服務人員被要求回答一個問題，如圖 1 所示。每次觀察 10 次。最終的決定是否會禮貌的問"我可以幫你嗎？"每次觀察都是通過投票方式進行的。我們讓人們在鼎泰豐服務，每人三票，其他人一票一票。精確度結果如 Table 4 所示，最正確的標籤是由被觀察者給出。圖五 顯示的是人與機器人在分辨被觀察者需不需要幫忙的圓餅分布圖。

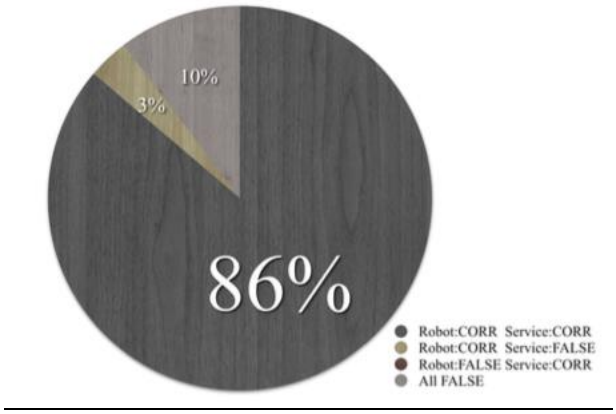

圖5

	Social Co-Robot	People in Service Industry
Accuracy (%)	90%	86.67%
TP rate (%)	56.67%	53.33%
FP rate (%)	10%	10%
TN rate (%)	33.33%	33.33%
FN rate (%)	0%	3.33%

Table 4

產業應用及其重要性

在本研究中，我們展示了一個令人興奮的結果，機器人具有潛在的能力從情境環境中學習，以提供"正確的服務"。我們開發的最先進的技術，允許機器人成功地從順序的人類行為學習。因此，機器人可以提供更大的適當的服務，更友善和更體貼。在實驗結果中，我們發現人體語言揭示了人類思想的信息。此外，通過考慮時空因素，我們檢索了確定需要援助的重大進展。關於 HOG 特徵，它通過高斯樸素貝葉斯分類器得到 60%的準確度。但是，它適用於基於 LSTM 的分類器的 69%的準確性。在編碼圖像方面，SVM 分類器只能產生 49.5%的精度，僅僅是機會。然而，通過基於 LSTM 的分類器分類的順序編碼圖像的精度提高了 73%。在第三個實驗中，通過融合編碼圖像和 HOG 特徵，精度提高到 75%。第四個實驗表明，通過採用光流作為輸入特徵，CNNs 跟隨 LSTM 深度學習模型產生 78%的精度。令人鼓舞的結果表明，通過從觀察中學習，該機器人可以提供"正常服務"的重要一步。最後一個但不是最少的實驗，我們使實驗 4 中提到的深度學習模型在機器人上實時運行。然後，將機器人的預測與服務行業人員的決策進行比較。有趣的結果表明，社會共同機器人與服務業人士之間有 96%的一致意見。這個實驗表明，我們的深度學習方法可以具有與在服務行業工作的人相同的能力來感知他人的需求。