

語音及自然語言多模態偵謊辨識

Multimodal Deception Detection by Language and Natural Language

組別：B270

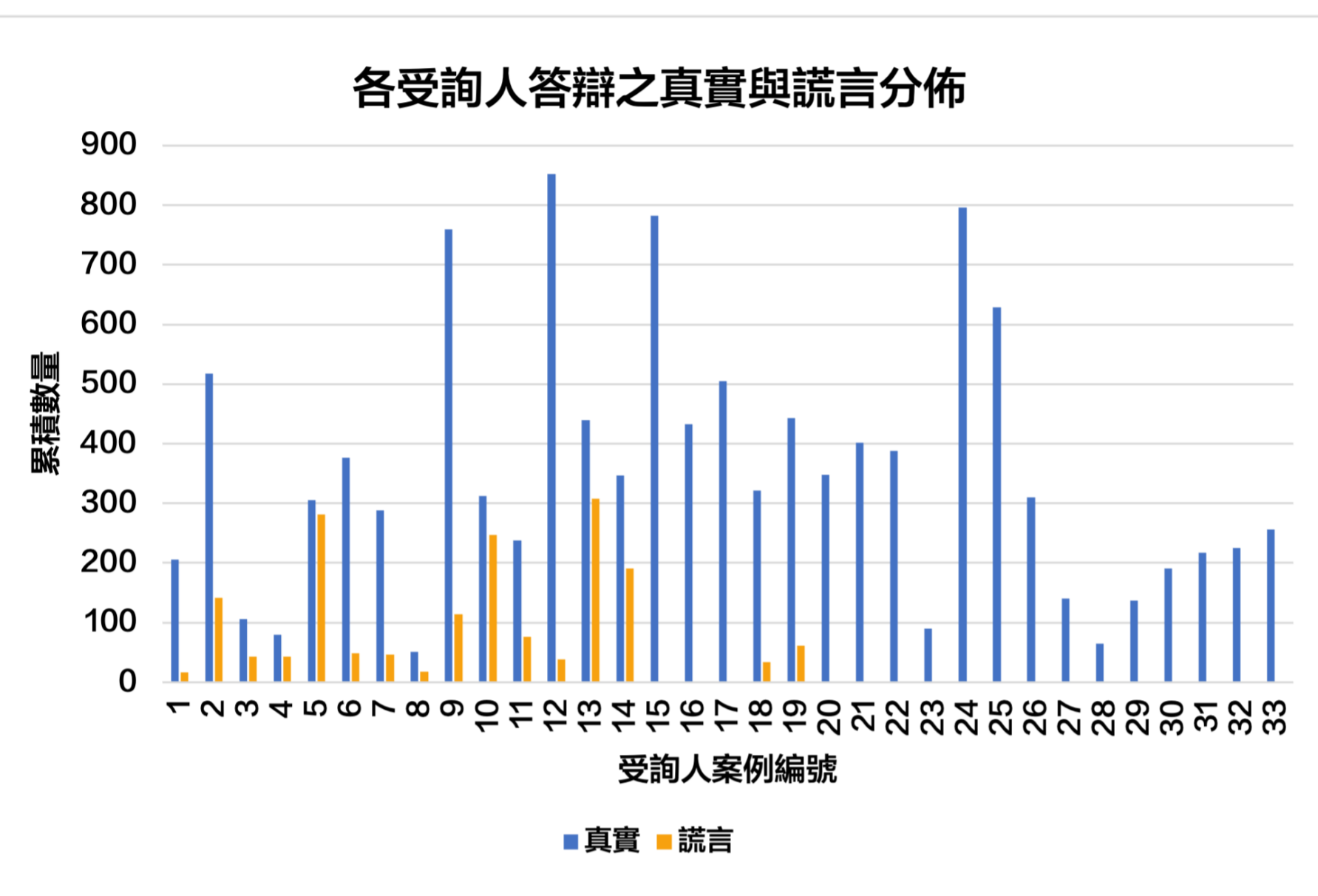
組員：王瑋毅、冉宜軒 指導教授：李祈均 教授

前言 Introduction

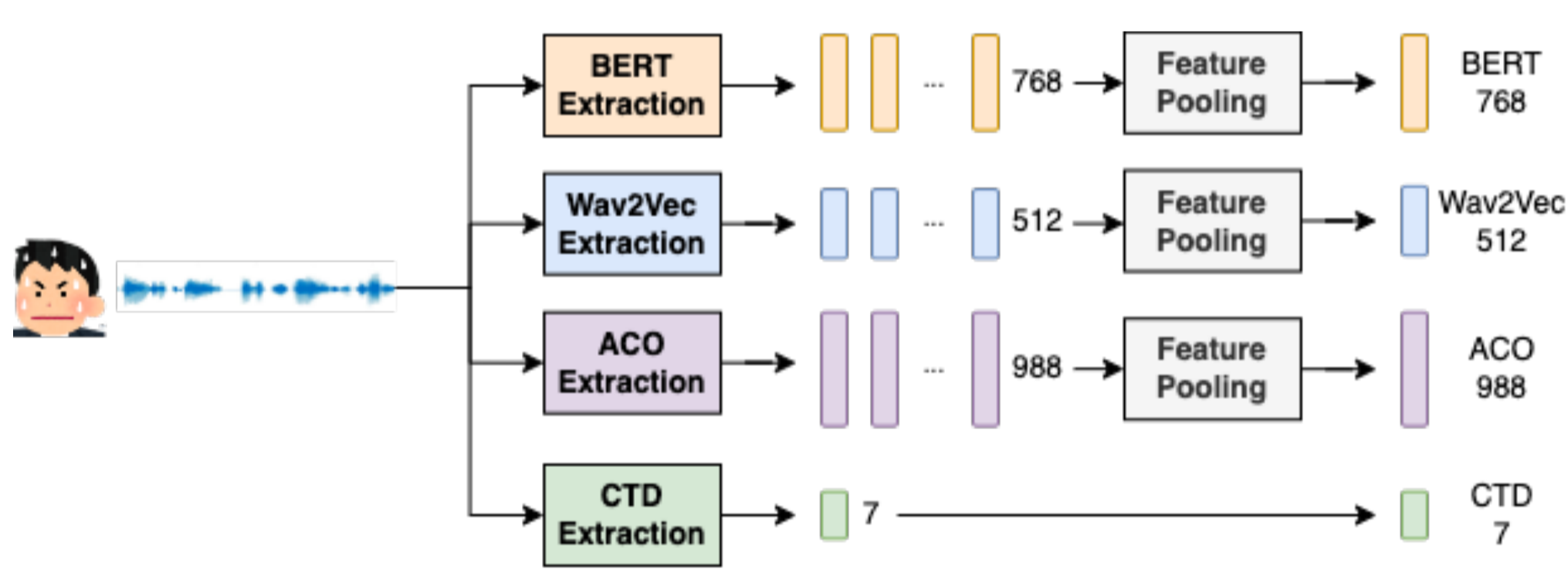
在法務部廉政署偵訊過程中，多數被認定為嫌疑人的受詢人時常以謊言扭曲關鍵事實，以躲避追緝，若是能判斷出受詢人哪些言論有較高的說謊可能性，則可以使廉政官將對談內容聚焦，減少偵訊時間，以加速辦案效率。本作利用偵訊時的「音訊」與「文本」紀錄來進行偵謊，透過人工智慧模型來判斷受詢人的口供內容是否屬實。

資料集 Dataset

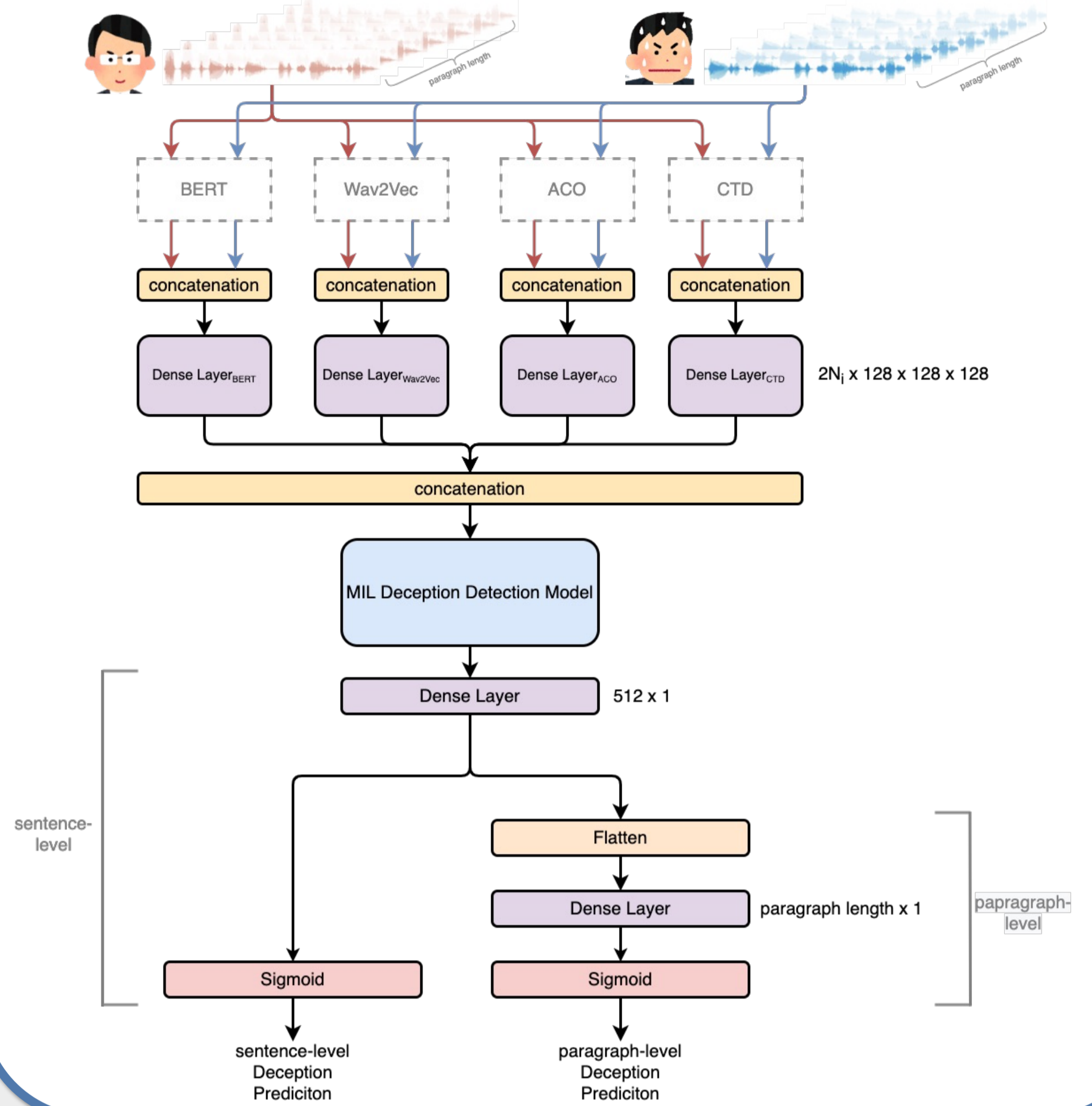
本作所使用之資料皆由法務部方面完成，一共收錄33則案例的對話紀錄，一共剪輯成 13265 筆的音檔，11558句實話 (87%)、1706句謊言 (13%)。



特徵萃取 Feature Extraction



模型架構 Framework



分析 Analysis

1. 使用多種特徵進行訓練可以提供模型更多資訊，相較於使用單一特徵時的 F1 score 總是在 0.5 附近左右徘徊，唯獨使用 BERT 時可以到 0.6，我們推測主因為人類說謊時，容易在句尾加上特定語助詞，而文本特徵較聲學特徵較易捕捉此種現象。
2. 使用對話(dialogue)類型的資料相較於僅使獨白(monologue)可以有效提升偵謊模型的表現，因為當受詢人意圖捏造謊言掩蓋事實時，容易使前後文有交代不清的現象，而此時廉政官會不自覺地增加提問頻率，而此種現象被 CTD 特徵所捕捉，所以廉政官的行為表現也透露了受詢人試圖說謊的行為。
3. 根據廉政官提問模式，我們所建立的多實例學習模型可以有效偵測受詢人在一主題段落的範圍內是否有說謊的跡象，其中主題段落的長度約為 15 句話，同時也提升針對逐句偵謊的表現，最佳的模型在主題段落層級的偵謊結果 F1 score 可以達到 0.73，而逐句層級的偵謊結果 F1 score 可以達到 0.72。

實驗結果 Results

model	features	feature pooling	paragraph size	conversation	sentence-level				paragraph-level			
					ACC	UAR	Precision	F1 score	ACC	UAR	Precision	F1 score
LSTM	BERT, ACO, CTD	mean	-	dialogue	0.7585	0.5844	0.5587	0.5636	-	-	-	-
LSTM	BERT, Wav2Vec	mean	-	dialogue	0.7835	0.5759	0.5621	0.5669	-	-	-	-
LSTM	BERT, Wav2Vec, CTD	mean	-	dialogue	0.7835	0.5759	0.5621	0.5669	-	-	-	-
Transformer	BERT, Wav2Vec, ACO, CTD	mean	-	dialogue	0.8379	0.5997	0.6191	0.6077	-	-	-	-
Transformer	BERT, ACO	mean	-	dialogue	0.8189	0.6277	0.6138	0.6199	-	-	-	-
Transformer	BERT, CTD	mean	-	dialogue	0.8122	0.6259	0.6195	0.6225	-	-	-	-
Transformer	BERT, ACO, CTD	mean	-	dialogue	0.8390	0.6290	0.6358	0.6322	-	-	-	-
Hopfield	BERT, ACO, CTD	mean	5	dialogue	0.7916	0.7480	0.6457	0.6655	0.5533	0.6448	0.5720	0.4991
Hopfield	BERT, ACO, CTD	mean	10	dialogue	0.8163	0.7747	0.6701	0.6958	0.8203	0.7450	0.6942	0.7127
Hopfield	BERT, ACO, CTD	mean	15	dialogue	0.8810	0.7163	0.7357	0.7253	0.8633	0.7072	0.7775	0.7332
Hopfield	BERT, ACO, CTD	mean	20	dialogue	0.8294	0.7839	0.6810	0.7089	0.6390	0.6237	0.5820	0.5696

參考文獻 Reference

1. Huang-Cheng Chou, Yi-Wen Liu, Chi-Chun Lee. Automatic Deception Detection using Multiple Speech and Language Communicative Descriptors in Dialogs. Cambridge University Press in association with Asia Pacific Signal and Information Processing Association, vol. 10, e5, 1-9.
2. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of NAACL- HLT 2019, 4171-4186.
3. Alexei Baevski, Henry Zhou, Abdelrahman Mohamed, Michael Auli. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. Facebook AI. (2020.06). [8] Ashish Vaswani, Noam Shazeer et al., Attention Is All You Need, 31st Conference on Neural Information Processing Systems (NIPS 2017).
4. Hubert Ramsauer, Bernhard Schöfl, Johannes Lehner et al. (2021) Hopfield Networks is All You Need.