

# 醫病訊息決策與對話語料分析

109753205 賴建郡、109753203蕭郁君、109753106王中伶

# CONTENTS

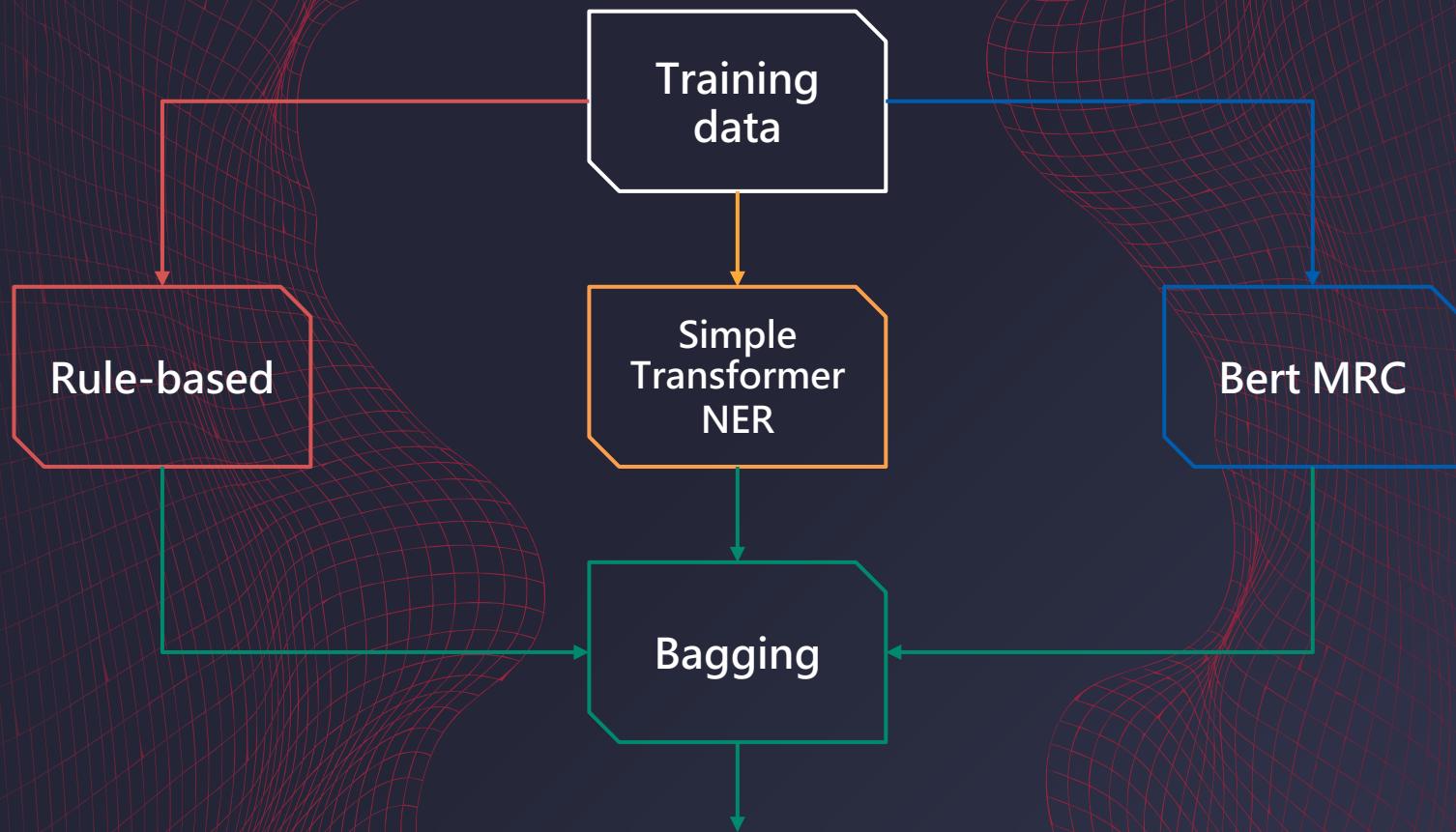
- 01 系統架構
- 02 系統解說
- 03 結果分析

The background features a dark blue gradient with two prominent, wavy red grid patterns. One grid is on the left, curving upwards and outwards, while the other is on the right, curving downwards and outwards. A thin white vertical line runs parallel to the right edge of the slide.

# 01

# 系統架構

# 系統架構



# 系統架構

## Rule-based

- 明確特徵

## Simple Transformer NER

- 各類別分開訓練
- 全部類別訓練

## Bert MRC

- 文意理解模型



## 02

# 系統解說

family profession id

contact money

time

Rule-based

## Rule-based

bagging

stop words

result

location

name

money

family

time

education

profession

轉簡體中文

Simple  
Transformer  
NER

轉繁體中文

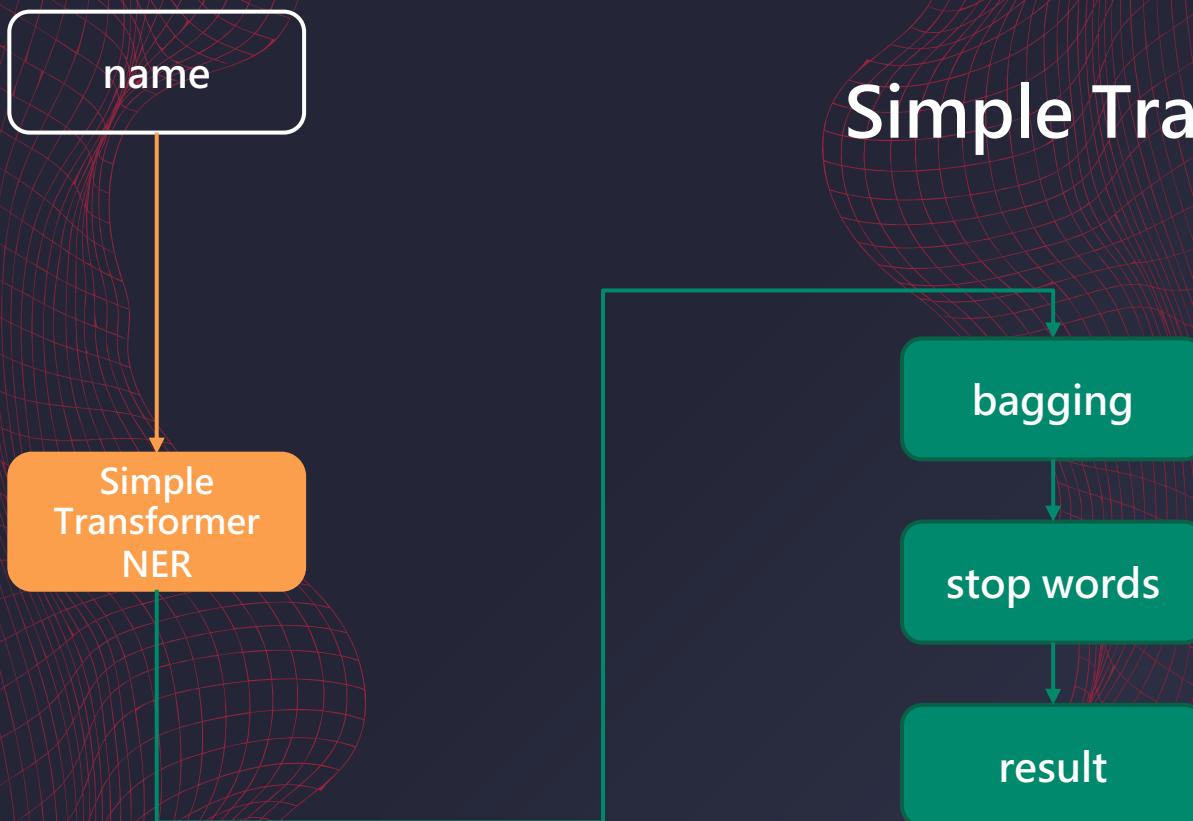
# Simple Transformer NER

bagging

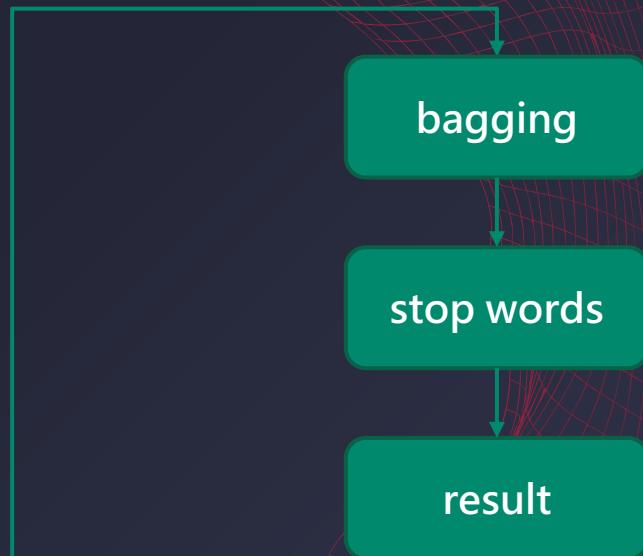
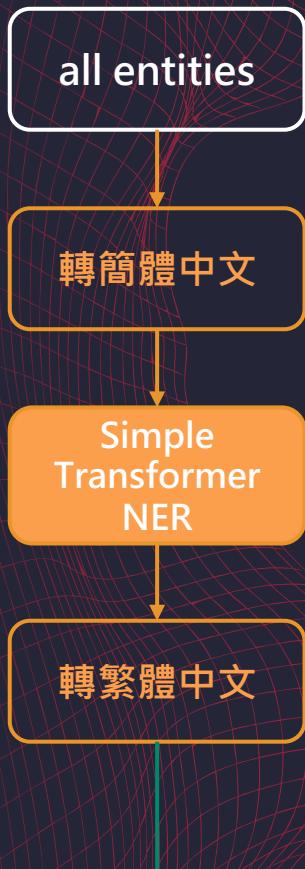
stop words

result

# Simple Transformer NER



# Simple Transformer NER

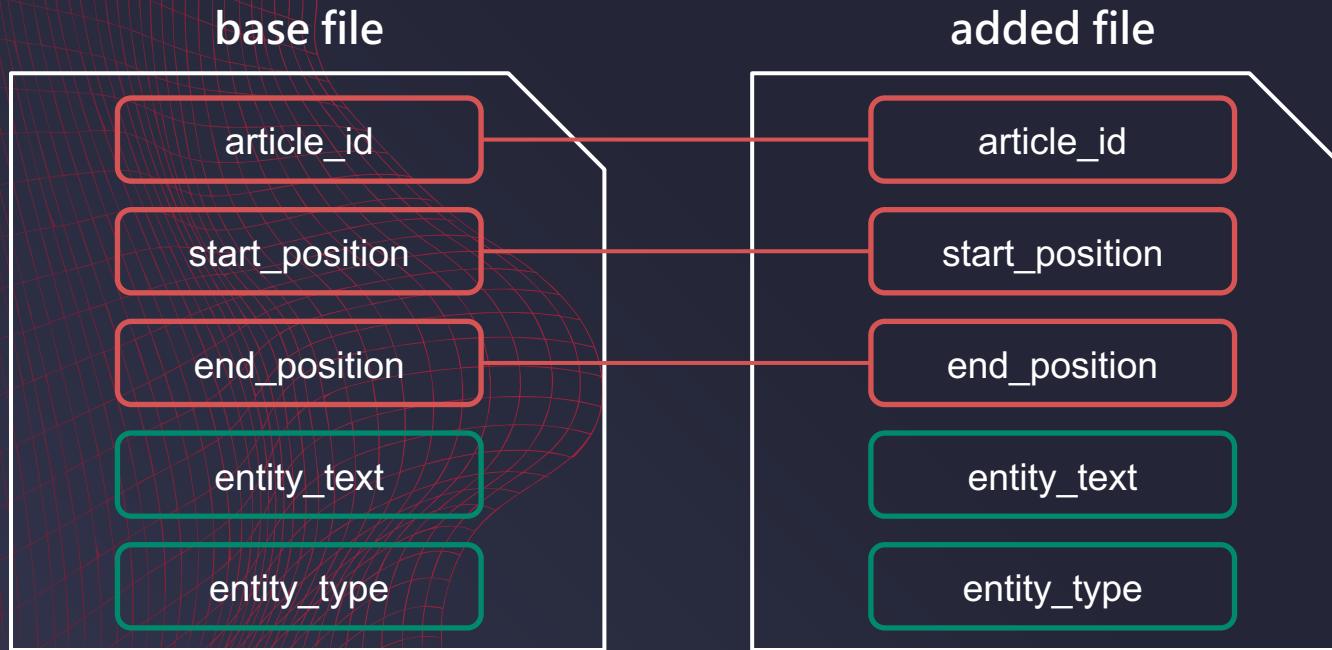


# Bert MRC



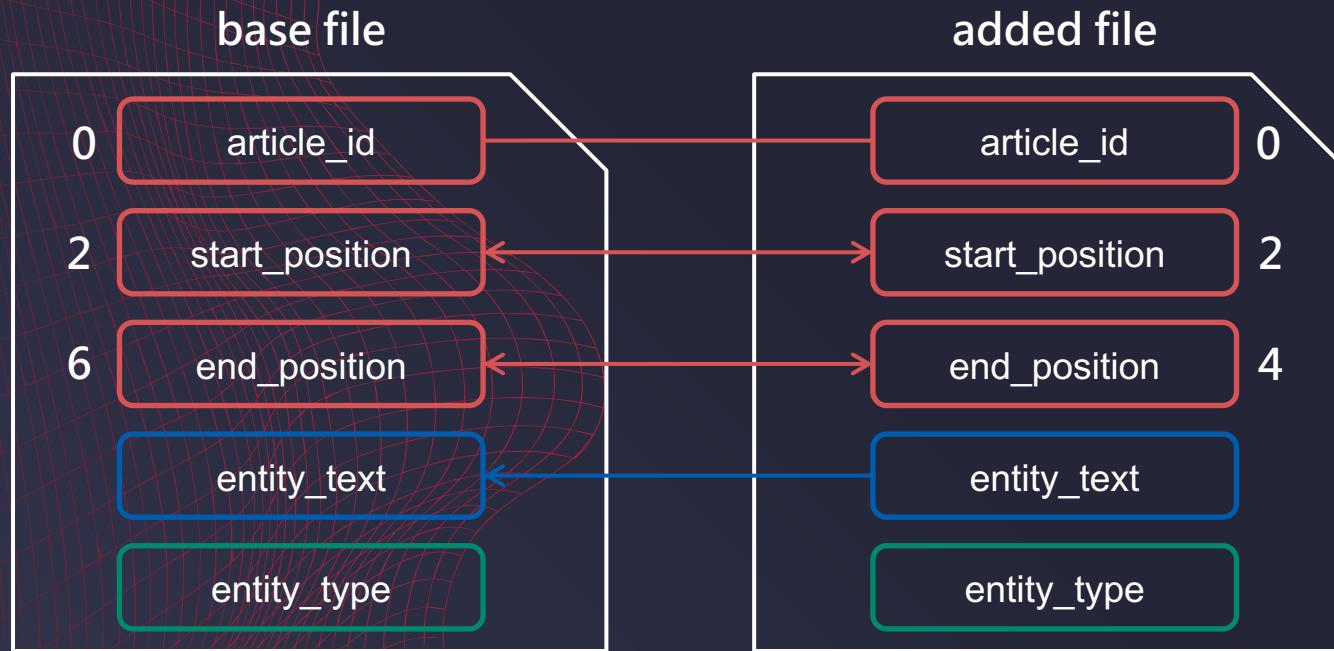
# Bagging

- 依照表現高到低之次序合併
- 清除相同或重疊特徵



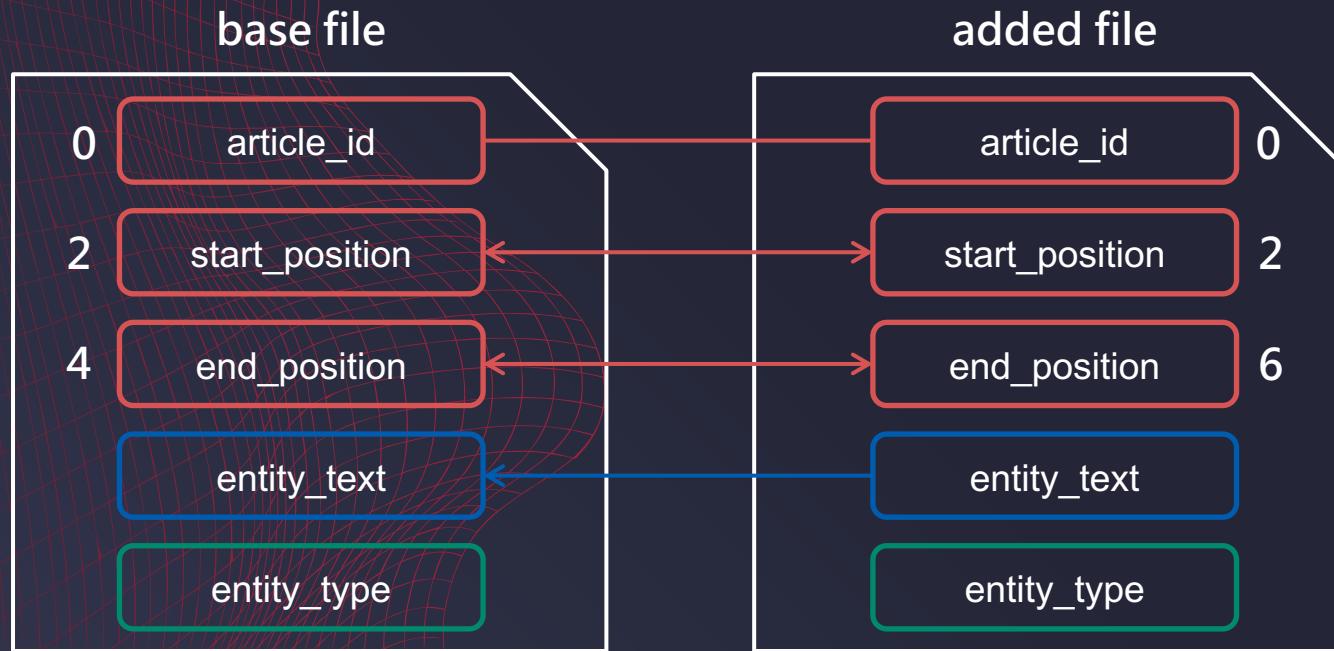
# Bagging

- 依照表現高到低之次序合併
- 清除相同或重疊特徵



# Bagging

- 依照表現高到低之次序合併
- 清除相同或重疊特徵



# Stop words

將不合理於某類別  
之字串清除



03

## 結果分析

---

# Public Leaderboard



Recall



F1-score



Precision

# Private Leaderboard

78.11%

F1-score

# 嘗試方法

- crf
- Bi-lstm+crf
- Simple Transformer  
NER
- bert-MRC

將訓練資料  
轉為簡體中文



- 攤平刪除相同特徵
- 排序過濾重疊特徵

# 詞頻分析

word	label_count	total_count	百分比
今天	171	450	38%
昨天	21	37	56%
明天	16	33	48%
後天	11	19	57%
前天	28	30	93%
早上	96	153	62%
中午	12	25	48%
晚上	27	105	25%
凌晨	1	1	100%

Thanks for Listening

---