

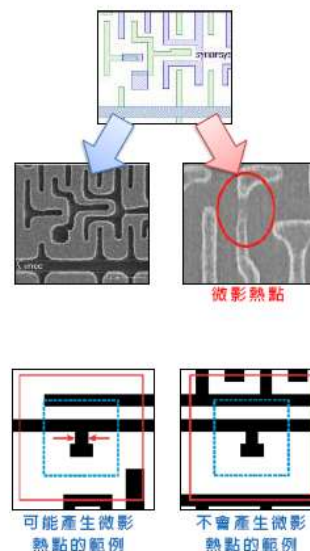
# Term Project

- The term project is carried out in teams of 1-2 students
- One-page project proposal (due date **May 17**, 2021)
  - The problem description
    - Background, previous work, motivation
  - Methods
    - Assumptions, proposed flowchart or steps
  - Possible results
  - References
    - Books, papers, websites
- Project demo & presentation (**June 3, 7, 10, 17**)
- Final report (**June 28**)
  - The final report (4-8 pages) should extend your proposal to include detailed methods, results and discussion

## Layout Pattern Classification for Hotspot Detection

100062527 林聖淵、100062523 陳靜怡

- 動機：
  - 電路尺寸下降 + 光微影製程誤差相對大
  - 部份路徑易產生 **微影熱點(lithography hotspot)**
    - 希望在電路佈局階段，就把可能產生微影熱點的電路偵測出
- 方法：
  - 資料：路徑位置、輪廓、骨架
  - 分類：
    - 透過LDA將原始資料進行降維
    - 透過SVM、Artificial Neural Network、KNN進行分類
- 實驗 + 分析：
  - Precision and Recall
  - SVM設定不同參數的實驗結果
  - 不同LDA維度影響辨識率的討論



## Correct Preposition Errors

9965703 張至、100062703 張堯

- 動機：
  - 使用者輸入可能含錯誤介係詞的英文句子  
系統輸出正確句子
  - 不易取得錯誤的句子當作訓練資料

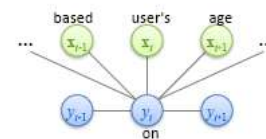
*On this paper, we present a new classifier based user's age.*



輸出

*In this paper, we present a new classifier based **on** user's age.*

- 方法：
  - 正確 + 錯誤句全改為中介形式
    - 無介係詞
  - Conditional Random Field (CRF)訓練model
    - Context-dependent Classification (Ch9)



- 實驗 + 分析：
  - 資料：British National Corpus
    - 包含13個常用介係詞、共約400萬字
  - 計算準確率 + 針對易誤用介係詞進行討論



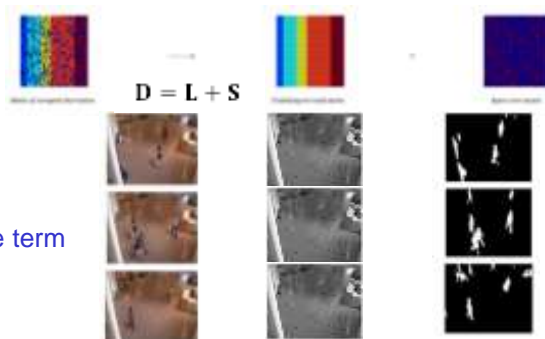
## Moving Object Detection with Robust PCA

101062547 黃啟清  
101062576 劉品均

- 動機
  - 將移動物體偵測視為image decomposition 問題

$$\min_{L, S} \text{rank}(L) + \lambda \|S\|_0$$

$$s. t. D = L + S$$



- 方法
  - 實作與比較三種方法
  - Robust PCA
  - Robust PCA with noise term
  - Multi-task robust PCA

- 實驗與分析
  - 對sparse term套入不同的threshold作二值化，以F-measure找出最適合的threshold
  - 套入不同threshold計算precision以及recall，並畫出ROC curve作為評量

101065505 林孟樺  
101062630 饒彥章

- Training Data: 400 個音檔 (200 個電子合成語音音檔+200 個人聲語音音檔)
  - Testing Data 1: 400 個音檔 (200 個電子合成語音音檔+200 個人聲語音音檔)
  - Testing Data 2: 400 個音檔 (200 個實際電子合成語音音檔+200 個人聲語音音檔)
- 
- | Condition                     | frame-based (%) | segment-based (%) |
|-------------------------------|-----------------|-------------------|
| without background music      | ~95             | ~90               |
| with background music         | ~85             | ~80               |
| with background music + noise | ~75             | ~70               |
- 分析選取不同feature數的影響(recognition rate)
  - 比較frame-based及segment-based下的差異
  - 比較有/不加背景音樂的差異

9960106 陳立昂  
9860106 許峻榮

- [illegible]



# NBA Game Prediction

- 動機
  - 透過online learning的方式對每一場NBA比賽做賽前預測
- 方法
  - Predict personal performance
    - 使用過去的比賽數據做加權平均，時間上離現在越接近或是和這次對手相同的場次會有較高的weight
  - Predict team work
    - 以預測出的球員數據當作feature vector，
    - Regression: 預測兩隊分別會得幾分，並以此來判斷誰輸誰贏
    - SVM: 在每個feature分別預測結果(+1/-1)，總和較高的即為獲勝球隊
- 實驗與分析
  - Data: NBA官網 2010~2011、2011~2012、2012~2013三個賽季
  - 結果討論
    - 當球隊與上一季狀況相差不大時有不錯的效果
    - 每支球隊因為有不同的特性，而可能有各自適合的model

# 自動語音點餐系統

- 動機
  - 飲料店常因顧客思考點餐內容或複雜的點餐組合而使隊伍大排長龍
  - 設計出一個系統能自動將一句點餐轉為文字，節省點餐人力
    - Audio input: 呃嗯...烏龍...綠茶... ..半糖去冰... ..兩杯
    - Output: 烏龍綠茶 半糖 去冰 兩杯
- 方法
  - Acoustic model: 39 MFCC + HMM (5 hidden states)
    - 飲料名稱、甜度、冰塊、數量4個分類器中各有14、10、15、15類別
  - Language model: HTK (Hidden Markov Model toolkit)
    - 將聲母、韻母拼成一個字 & 將字與字拼成一個詞:
- 實驗與分析
 

	飲料名稱	冰塊	甜度	數量
整句準確度	72.96%	82.05%	74.36%	83.33%
單字準確度	81.70%	91.03%	87.18%	91.67%

  - 資料蒐集: 11位男性、7位女性，其中4位在背景嘈雜環境中錄音
  - 討論：中文字的發音重疊性高，像是「紅、龍、檬」、「微、無」、「去、全」、「兩、六」等，且台灣人講話多簡化咬字，「四、十」這種些微的差異有時連人耳都難以分辨

# Tic-Tac-Toe Game 井字遊戲

- 動機
  - To develop a game playing A.I. based on the AlphaGo's method
  - Tic-Tac-Toe is selected as the game to be played
- 方法
  - The program is developed based on the method used by AlphaGo
    - Two types of neural networks: policy network & value network
    - Search tree
- 實驗與分析
  - Training data of the policy network is generated with an algorithm based on existing perfect strategy

Opponent	random actions		perfect strategy	
	first	second	first	second
1Strategy loses...	987	866	0	0
wins	0	0	0	0
draws	11	125	1000	1000

Table 1: 1Strategy play against random actions and perfect strategy.

Turn 1:	Turn 2:	Turn 3:	Turn 4:																																														
<table><tr><td></td><td></td><td></td></tr><tr><td>x</td><td></td><td></td></tr><tr><td></td><td></td><td></td></tr></table>				x						<table><tr><td></td><td></td><td></td></tr><tr><td>x</td><td>o</td><td></td></tr><tr><td></td><td></td><td></td></tr></table>				x	o					<table><tr><td></td><td></td><td>x</td></tr><tr><td>x</td><td>o</td><td></td></tr><tr><td></td><td></td><td></td></tr></table>			x	x	o					<table><tr><td></td><td>o</td><td>x</td></tr><tr><td>x</td><td>o</td><td></td></tr><tr><td></td><td></td><td></td></tr></table>		o	x	x	o														
x																																																	
x	o																																																
		x																																															
x	o																																																
	o	x																																															
x	o																																																
Turn 5:	Turn 6:	Turn 7:	Turn 8:	Turn 9:																																													
<table><tr><td></td><td>o</td><td>x</td></tr><tr><td>x</td><td>o</td><td></td></tr><tr><td>x</td><td></td><td></td></tr></table>		o	x	x	o		x			<table><tr><td></td><td>o</td><td>x</td></tr><tr><td>x</td><td>o</td><td>o</td></tr><tr><td>x</td><td></td><td></td></tr></table>		o	x	x	o	o	x			<table><tr><td></td><td>o</td><td>x</td></tr><tr><td>x</td><td>o</td><td>o</td></tr><tr><td>x</td><td>x</td><td></td></tr></table>		o	x	x	o	o	x	x		<table><tr><td></td><td>o</td><td>x</td></tr><tr><td>x</td><td>o</td><td>o</td></tr><tr><td>x</td><td>x</td><td>o</td></tr></table>		o	x	x	o	o	x	x	o	<table><tr><td>x</td><td>o</td><td>x</td></tr><tr><td>x</td><td>o</td><td>o</td></tr><tr><td>x</td><td>x</td><td>o</td></tr></table>	x	o	x	x	o	o	x	x	o
	o	x																																															
x	o																																																
x																																																	
	o	x																																															
x	o	o																																															
x																																																	
	o	x																																															
x	o	o																																															
x	x																																																
	o	x																																															
x	o	o																																															
x	x	o																																															
x	o	x																																															
x	o	o																																															
x	x	o																																															

Figure 2: A9 game where 1Strategy lost. Player-1 uses random actions.

## PTT 酸文辨識

105065507 姚舜懷  
105062507 陳星宇  
105062525 林宏鑑

- 動機
  - PTT 上常有發文或是網友一言不合透過留言互酸
  - 如果可以辨識一篇發文或底下的酸言佔比多高，我們認為可以幫助鄉民更輕鬆選擇符合自己偏好的文章
- 方法
  - Data collection: 自行撰寫爬蟲抓取 PTT 站上文章資料
  - Features: TF-IDF, word2vec
  - Classifiers: Naïve Bayes, SVM, LSTM
- 實驗與分析
  - Data imbalance:酸文與中性文數量差距大

Model	Feature	Unl. acc. - Bk	Unl. acc. - Mh	Unl. acc. - Lg
Naive Bayes (GaussianNB)	word2vec	49.4%	51.0%	51.8%
MM	word2vec	55.2%	50.8%	52.1%
Naive Bayes (MultinomialNB)	TF-IDF	62.5%	68.5%	72.4%
SVM	TF-IDF	68.7%	68.7%	71.8%

Model	Feature	Unl. acc. - Bk	Unl. acc. - Mh	Unl. acc. - Lg
Naive Bayes (GaussianNB)	word2vec	49.4%	51.0%	51.8%
SVM	word2vec	55.2%	50.8%	52.1%
MM	word2vec	68.8%	72.3%	73.2%
MM	word2vec	64.5%	67.3%	70.1%

# 美國職棒大聯盟 (MLB)戰力分析

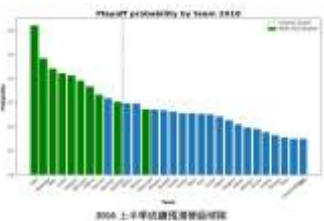
104034031 游新新

107061604 李杰倫

- 動機
  - 國外很多棒球網站每年開季前都會預測今年度能進到季後賽的球隊
  - 我們針對MLB中 30支球隊近10年來的數據進行分析，利用機器學習方法來看這些球隊花大錢增加戰力是否有實質幫助？又或者是哪些原因才是背後影響的關鍵？
- 方法
  - Data collection: Baseball reference
  - Feature selection: Wrapper method vs Filter method (52-D-> 9D)
  - Test on 5 classifiers
- 實驗與分析

SVM	Decision Tree	KNN	Adaboost	MLP
70%	80%	83%	80%	83%

利用 2017 作為訓練集 2018 上半年成績作為測試集



# 植物品種鑑定

107003815 鄭良加

- 動機
  - 植物種類多達450,000種，不易辨識
  - 希望透過手機拍照，並結合植物百科，以開發快速辨識模型
- 方法
  - Data collection: 8 public domain data sets
  - Feature extraction: PCA/LAD
  - Multiclass classification
- 實驗與分析



Dataset	Species	Sample Size
Flavia	33	60 / species
Swedish	15	75 / species
UCI	40	8-13 / species
Oxford flower	17	80 / species
CR Leaves	102	40-250 / species
PlantCLEF 2013 (Natural)	28	255 / species
PlantCLEF 2013 (Artificial)	66	9607 (train) 1194 (test)
PlantCLEF 2013 (Natural)	57	2585 (train) 525 (test)
Tree Canopy	8	890-26255 / species

- 針對課堂上所教之6種不同降維方式(PCA, LDA, ...)降到2維做一觀察，並加入 t-SNE 方法(希望樣本間在高維對應的高斯核函數相似度同樣在低維可以被保留)



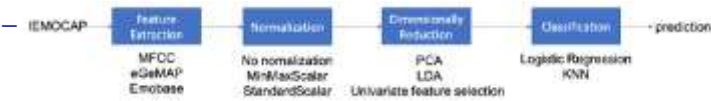
# 語音情緒分類模型

107062705 陳家昕  
107003817 張崇智

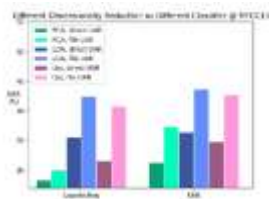
- 問題描述

- 全球智慧家庭裝置中以智慧音箱成長最為顯著[1]，語音已成為與智慧終端互動不可或缺的方式，因此擬針對聊天時常出現的 4 類情緒(喜、怒、哀、中性)進行辨識，在分類出使用者的情緒後，智慧終端即可依據狀況給予適當的回應。

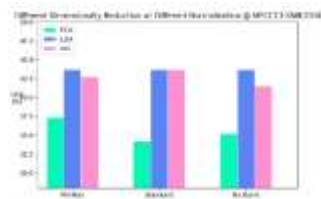
- 方法



- 實驗與分析



[figure 4]在 MFCC13 下實驗不同降維與分類器的組合，UAR 越高越好。



[figure 5]在 MFCC13 下實驗不同正規化方法配上 LDA(N=3)和 KNN(K=150)，UAR 越高越好。