

Portfolio - Data Analysis

Applicant: Chih-Yen Ou

Outline



- **Master Thesis**
 - **Speech Recognition System**
 - **Case Classification**



- **Text Analysis**



- **Sentiment Analysis**



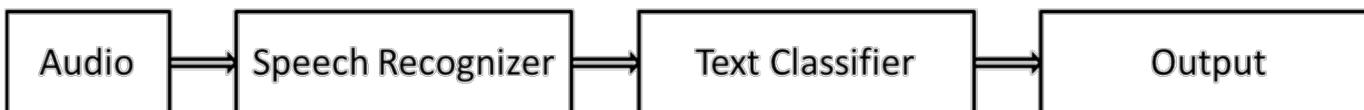


Master Thesis



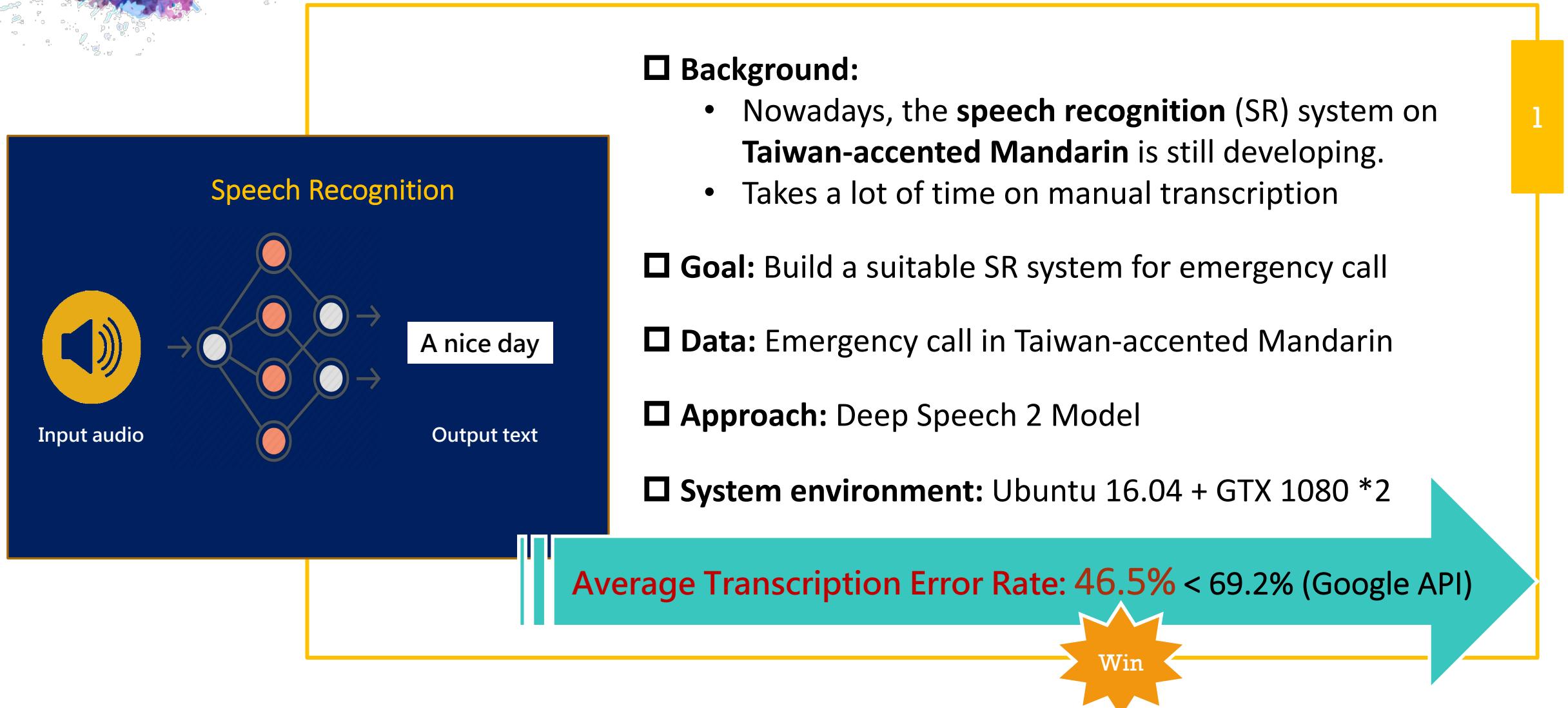
• Master Thesis

- **Title:** Natural Language Processing (NLP) based Analysis of the Emergency Medical Service Dispatch Communication: A Case Study of Major Trauma
- **Goal:**
 1. Build an automatic severity recognition system to assist ambulance dispatchers to recognize the injury severity
 2. Identify representative words for major trauma cases
- **Data:** Emergency call of car accident case (Taipei, 2018)
- **Approach:** Deep Speech 2, Support Vector Machine (Use python programming)
- **Flowchart:** classify call content into major trauma and non-major trauma case



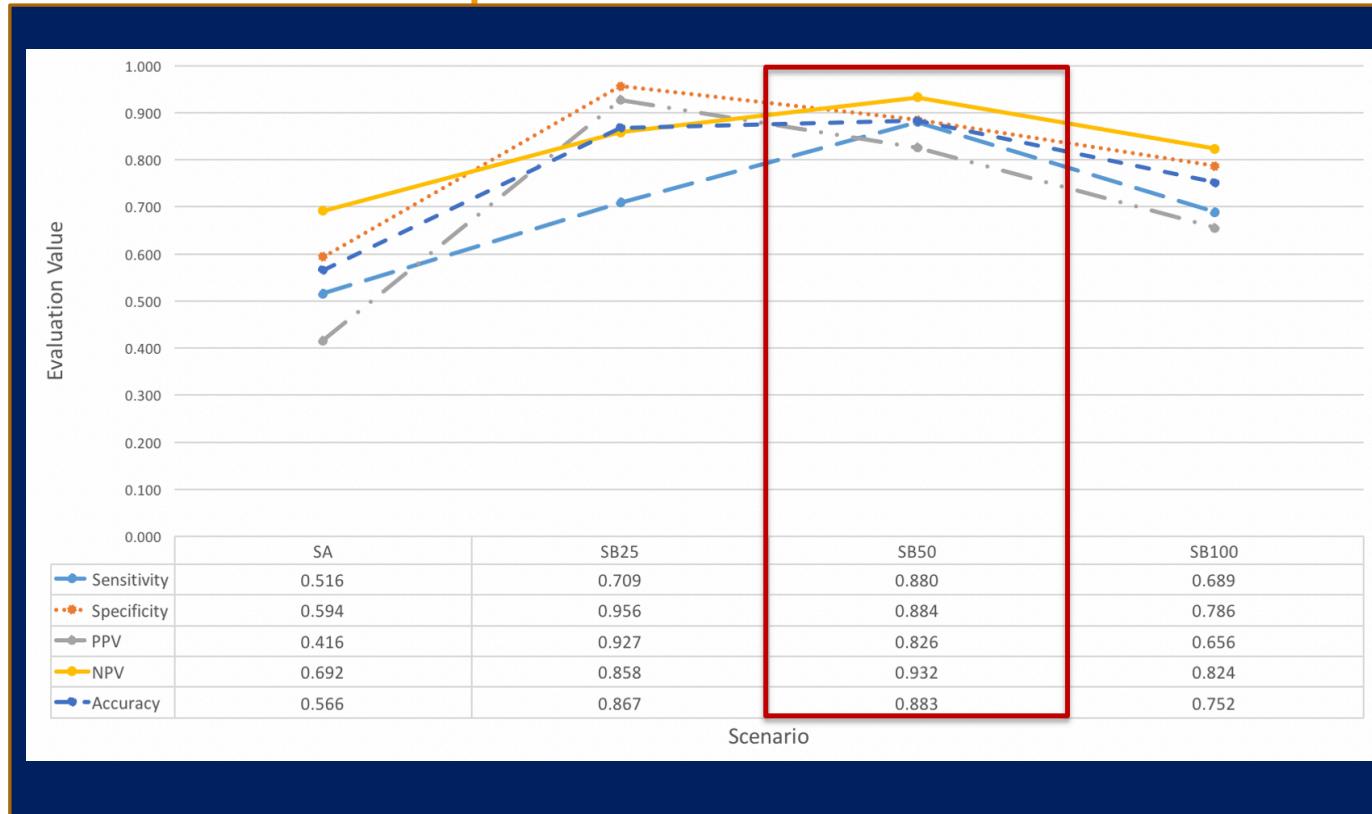


• Speech Recognition System





• Text Analysis – Case Classification Result



□ Background:

- Ambulance dispatcher **faces the difficulty of recognizing major trauma** on emergency call

□ Approach: SVM, term frequency-inverse document frequency (TF-IDF)

□ Representative words:

- major trauma: breath, hurry
- non-major trauma: awake, needed

*Sensitivity (recall)

*Specificity (recall of negative samples)

*Positive Predictive Value (precision)

*Negative Predictive Value (precision of negative samples)



Text **Analysis**



• Text Analysis—Educational News

15 common words on each category



Taiwan



Hong Kong

- **Introduction:** Since I was interested in reading educational news, I was curious about the difference of them between Taiwan and Hong Kong. Thus, I used web crawling to collect the articles and **analyzed the common words**.
- **Data:** The News Lens - Taiwan (264 articles) , Hong Kong (101 articles)
- **Data preprocessing:** remove punctuation, HTML tags, English words
- **Result:** The most common words in Taiwan and Hong Kong's articles are similar. Besides, the highest word similarity are “children” and “parents.”



Sentiment Analysis



• Sentiment Analysis on Movie Review

		predicted label Positive	Negative
true label	Positive	6491	1167
	Negative	1001	6341

Result:

- Precision: 0.845
- Recall: 0.864
- Accuracy: 0.856

□ Introduction:

- Some people write movie review without giving a score.
- Build a classifier to **predict the sentiment of reviews** (positive or negative).

□ Dataset: [IMDB Review Dataset](#)

- Training data: 35000
- Testing data: 15000

□ Data preprocessing: delete special characters, check the word length, transform upper case into lower case

□ Approach: Naïve Bayes



• Sentiment Analysis on Movie Review

Most popular Movie (2018)



(*word in red: wrong classification)

Review	Positive?
"Black Panther" lived up to the hype.	No
It's uniquely satisfying popcorn entertainment, a triumphant trifecta of subject matter, talent and timing.	Yes
When it comes to creative visuals, engaging action and likable characters, "Black Panther" stands confidently next to the best fare offered up by the Marvel Cinematic Universe.	Yes



• Sentiment Analysis on Movie Review

(*word in red: wrong classification)

Box Office Failure (2018)

ROBIN HOOD OFFICIAL TRAILER

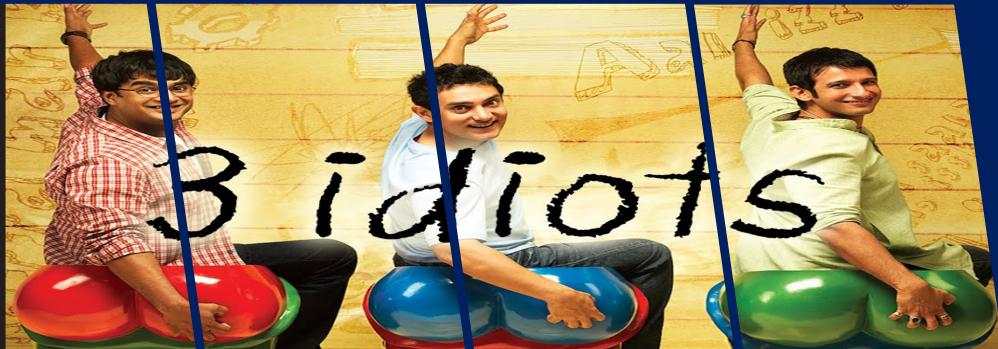
Review	Positive?
I wanted to like it. I couldn't manage it.	No
Open minded and was excited...then very disappointed	No
The Death of a Legend	Yes



• Sentiment Analysis on Movie Review

Blockbuster

HD JUKEBOX



Review	Positive?
Wonderfully entertaining.	Yes
easily the best Indian movie in 2009.... and one of the best ever	Yes
idiots is a masterpiece and easily one of the greatest Bollywood movies of all time	Yes



thank
you