

The Technologies We Used (or Not)
in
Formosa Grand Challenge

Archer Tsai 蔡幸育

Intellection, daVinci Lab/AICS Center, ASUS

Our Team

Speech Team & NLP Team of ASUS daVinci Lab (2010~2019)

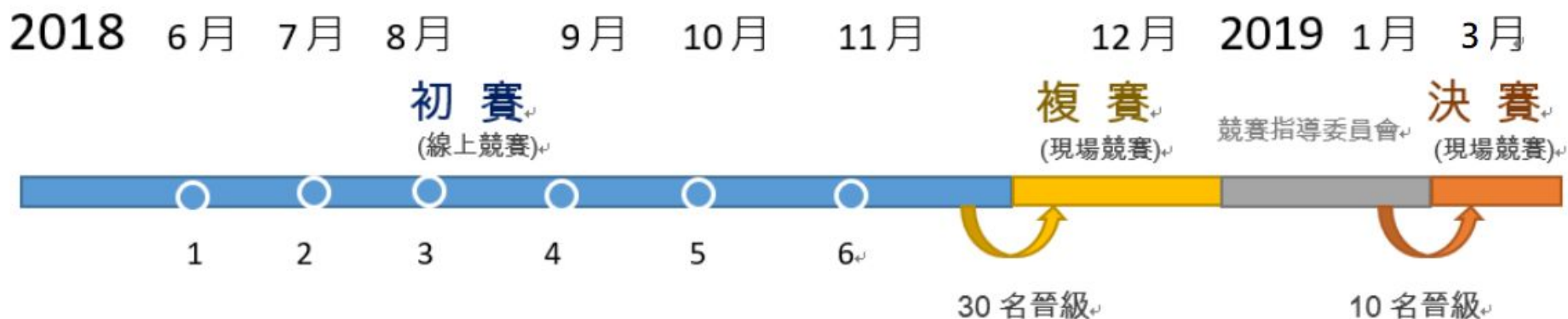


AICS Center (2019~)

- ASUS Intelligent Cloud Services (AICS)
- Intern Program: <http://ehr.asus.com/InternProgram.htm>
- Summer Intern
 - 2 month full-time(July ~ Aug)
 - \$42000/Month
- Long-Term Intern
 - 10 Hrs/week



Roadmap of Formosa AI Grand Challenge



Preparing from February to June

Involving Speech Recognition and Natural Language Processing

2019科技大擂台-與AI對話2(熱身賽)

報名時間:2019-10-01 ~ 2019-10-31

Outline

- Basic Concepts
 - Speech Recognition
 - Natural Language Processing

- Formosa Grand Challenge: Talk to AI
 - Challenges of Speech Recognition
 - NLP Solutions
 - Short Answering Questions

Basic Concepts

Speech Recognition

- 3rd Party Solution (Google, iFlyTek...)
 - Show me the money!!
- End-to-end Model
 - Speech-to-Text
 - Speech-to-Text + External LM
 - [Towards End-to-End Speech Recognition - ISCSLP 2018](#)
- Conventional Method (Kaldi, Sphinx...)

Speech Recognition

- 3rd Party Solution (Google, iFlyTek...)
- End-to-end Model
- Conventional Method (Kaldi, Sphinx...)
 - Feature Extraction
 - Acoustic Model
 - Lexicon
 - Language Model
 - (RNN) LM Rescoring

Speech Recognition

- Conventional Method (Kaldi, Sphinx...)
 - Feature Extraction
 - MFCC、PNCC、PLP

Speech Recognition

- Conventional Method (Kaldi, Sphinx...)
 - Acoustic Model
 - Which Architecture?
 - CNN, TDNN, RNN...
 - What to Model:
 - Mono-phone, Bi-phone, Tri-phone, Syllable...
 - Phone/Syllable Set?
 - Tonal? Non-Tonal?
 - How many hidden layers?
 - How many nodes in each layer?

Speech Recognition

- Conventional Method (Kaldi, Sphinx...)
 - Lexicon
 - Mapping phone sequence to word
 - How many words should be able to be recognized?

字典	ㄓㄧˋ ㄉㄧㄢ	ㄓㄧˋ ㄉㄧㄢˇ	z i i d i a n	z i i 4 d i a n 3
----	---------	----------	---------------	-------------------

Speech Recognition

- Conventional Method (Kaldi, Sphinx...)
 - Language Model
 - Modeling the relationship among words
 - 到淡水[起碼/騎馬]要兩個小時
 - N-gram model
 - N?
 - Target speech?
 - How much training data required?
 - RNN model
 - LSTM

2-gram

到 淡水
淡水 騎馬
淡水 起碼
騎馬 要
起碼 要

3-gram

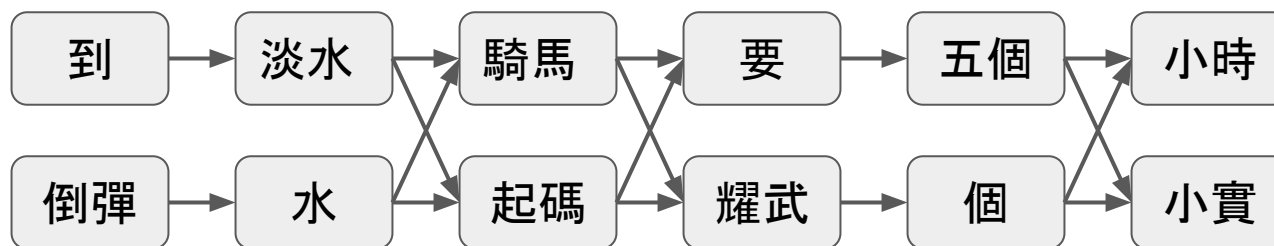
到 淡水 騎馬
到 淡水 起碼
淡水 騎馬 要
淡水 起碼 要

4-gram

到 淡水 騎馬 要
到 淡水 起碼 要

Speech Recognition

- Conventional Method (Kaldi, Sphinx...)
 - (RNN) LM Rescoring
 - Use a small LM to create lattice
 - Use a big LM to find a better path in the lattice



Lattice:

A lattice is a representation of the alternative word-sequences that are "sufficiently likely" for a particular utterance.

<http://www.kaldi-asr.org/doc/lattices.html>

Speech Recognition HOWTO

- Data Source
 - Labeled speech data
 - Buy Corpus
 - Youtube (or other video source with subtitle)
 - Prepare transcription and record
 - Collect data and label
 - Text data
 - Buy Text Corpus / Text Database
 - Collect from Internet

Speech Recognition HOWTO

- Data Preparation
 - Labeled speech data
 - Augmentation (Speed up/down, Mix noise signal...)
 - Text data
 - Decide the word set to be recognized
 - Generate Lexicon for the word set
 - Text Corpus
 - Cleaning data
 - Word Segmentation

Natural Language Processing

BERT

Natural Language Processing

- Different data labels are required for different purpose
 - Word Segmentation
 - Name Entity
 - Part of Speech
 - QA pair

Formosa Grand Challenge: Talk to AI

Speech Recognition

- 題型：華語文能力測驗題庫
 - 怎麼判斷題目敘述、問題與選項分界？
 - Ding Sound?
 - > Dry run時發現已經分開在不同檔案
 - 怎麼分辨語者是男是女
 - 這位小姐的意思是？
 - > Speaker Diarization
 - 怎麼正確辨識選項編號？
 - 依依？餓餓？姍姍？試試？
 - > 特製Language Model

Speaker Diarization

- Who speak when?
- I-Vector, X-Vector...
- Performance Issue
 - Use Gender detection instead?
 - Most questions could be solved without gender info -> skip

Option Extraction

- 缺乏LM訓練資料
- 選項編號與內文無邏輯關係

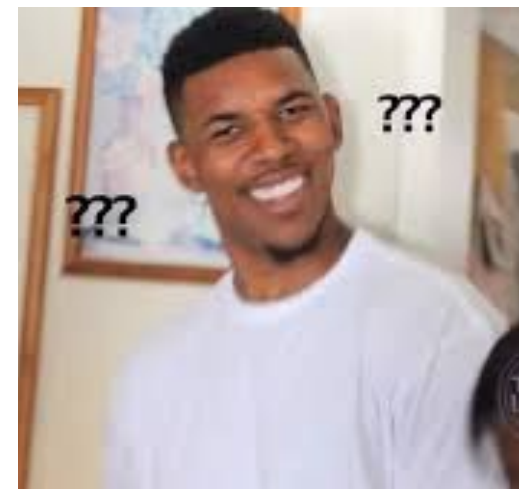
Option Extraction

- 做一個只包含選項編號的LM, 再做Interpolation
 - 只有1, 2, 3, 4
 - `<s>1</s>`
 - `<s>2</s>`
 - ...
 - 不影響原本一、二、三、四的機率

- 用Rule修正無法正確辨識的部分

Language Model

- OOV - 多為名詞, 只要求前後文一致
 - 雙吻前口蝠鱝?
 - 銀首闊腳袋鼯?
 - 盤定鋼索?
- Multi-Lingual - 加入常用英文字
 - 英語、台語、客語、原住民語
- 新聞、警廣新聞、古文 - 認命爬文吧



Background Noise

- 警廣新聞 現場收音片段
 - 不影響作答, 有多少是多少
- 環境音 (風聲、籃球場、工廠)
 - AM輕鬆處理
- 人聲噪音
 - Single Channel Cocktail Party Problem?
 - 用訊號處理方式, 降低人聲干擾 - 須重做AM

Models

- AM
 - CNN + TDNNF

- LM
 - 3-gram LM
 - RNNLM

Performance

	題目敘述 (句數 / CER)			問題	選項
	警廣	古文	其他		
預賽1	-	-	1500 / 7.44	5.41	9.21
預賽2	572 / 19.24	-	928 / 7.69	7.24	10.55
預賽3	600 / 11.55	900 / 25.97	-	6.52	10.73
預賽4	-	1500 / 10.72	-	7.69	10.9
預賽5	600 / -	900 / 4.45	-	4.09	7.3
預賽6	500 / -	1000 / 12.39	-	11.95	19.38

※預賽3開始，主辦單位賽後有提供題目文稿，但不包含警廣新聞現場收音部份

※預賽5、6時，未針對警廣新聞額外投入人力標注，故無準確率資料

NLP

- 訓練語料
 - 華語文測驗 3XX題
 - What else?

- SNLI, Multi-NLI + Google Translation
 - 不完美, 但是堪用

Decomposable Attention Model



陳黃威 Will

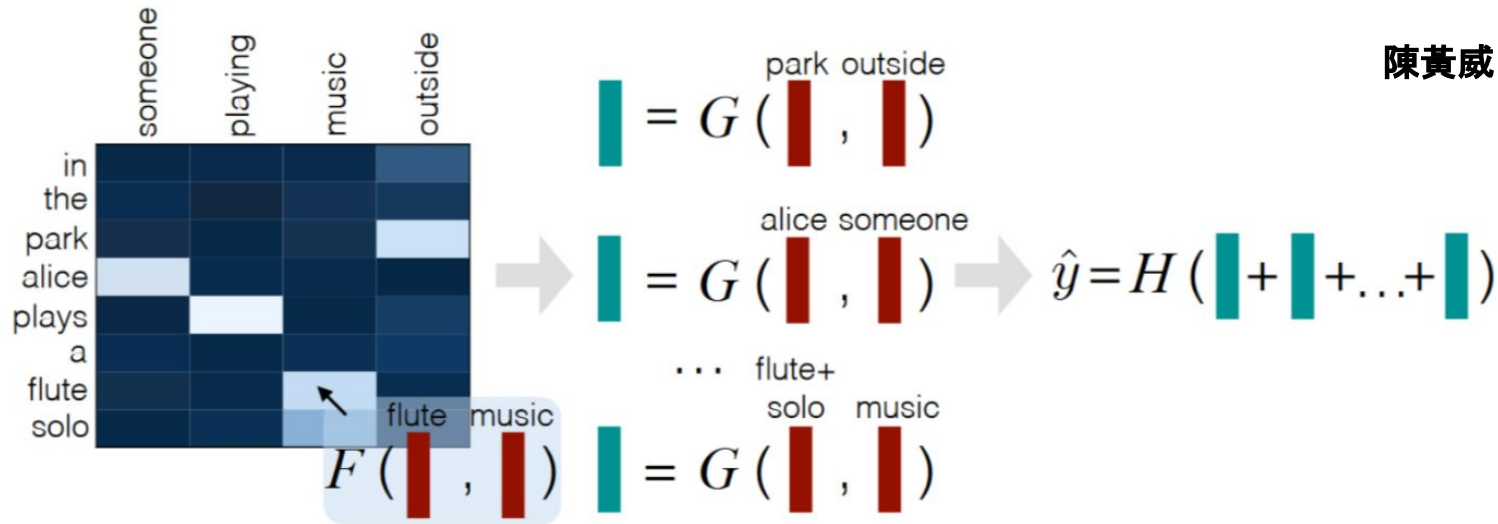
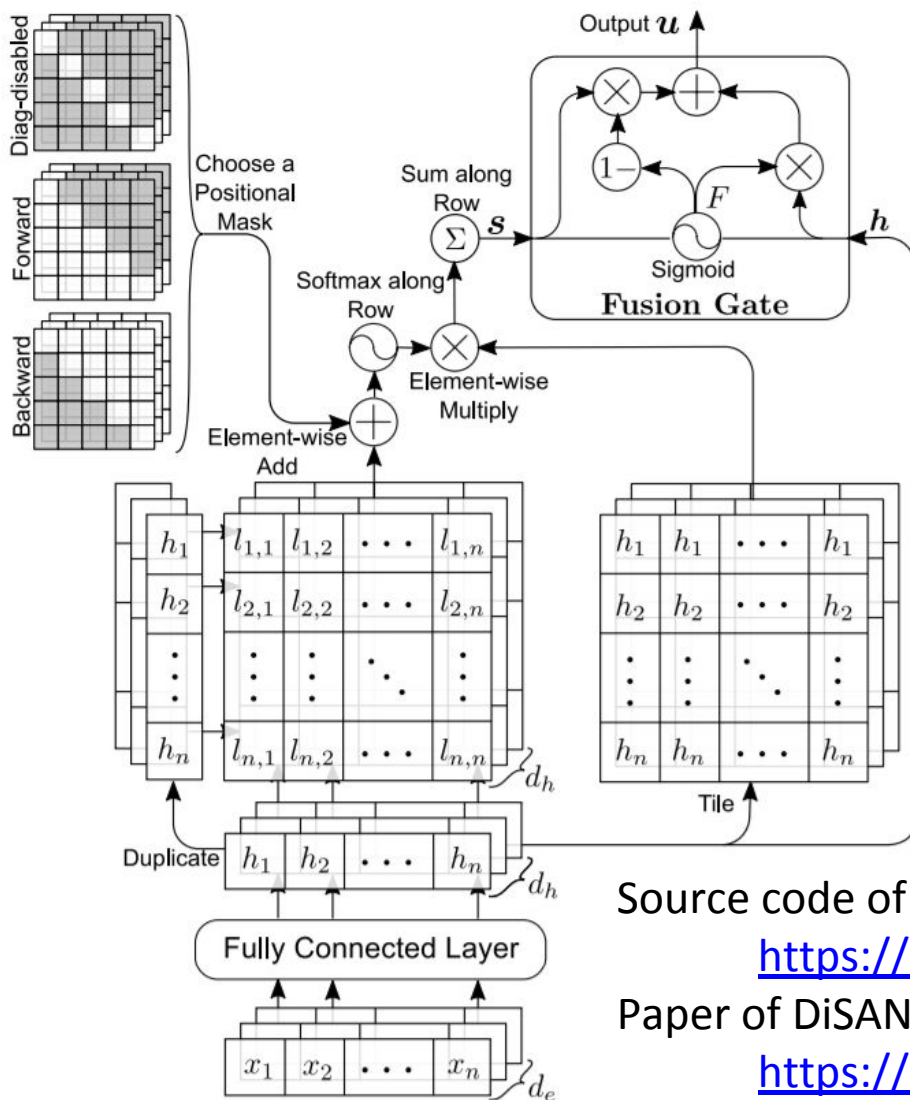


Figure 1: Pictorial overview of the approach, showing the *Attend* (left), *Compare* (center) and *Aggregate* (right) steps.

*A Decomposable Attention Model for Natural Language Inference, <https://arxiv.org/pdf/1606.01933.pdf>

Directional Self-Attention Network (DiSAN)



劉玠均

Source code of DiSAN:

<https://github.com/taoshen58/DiSAN>

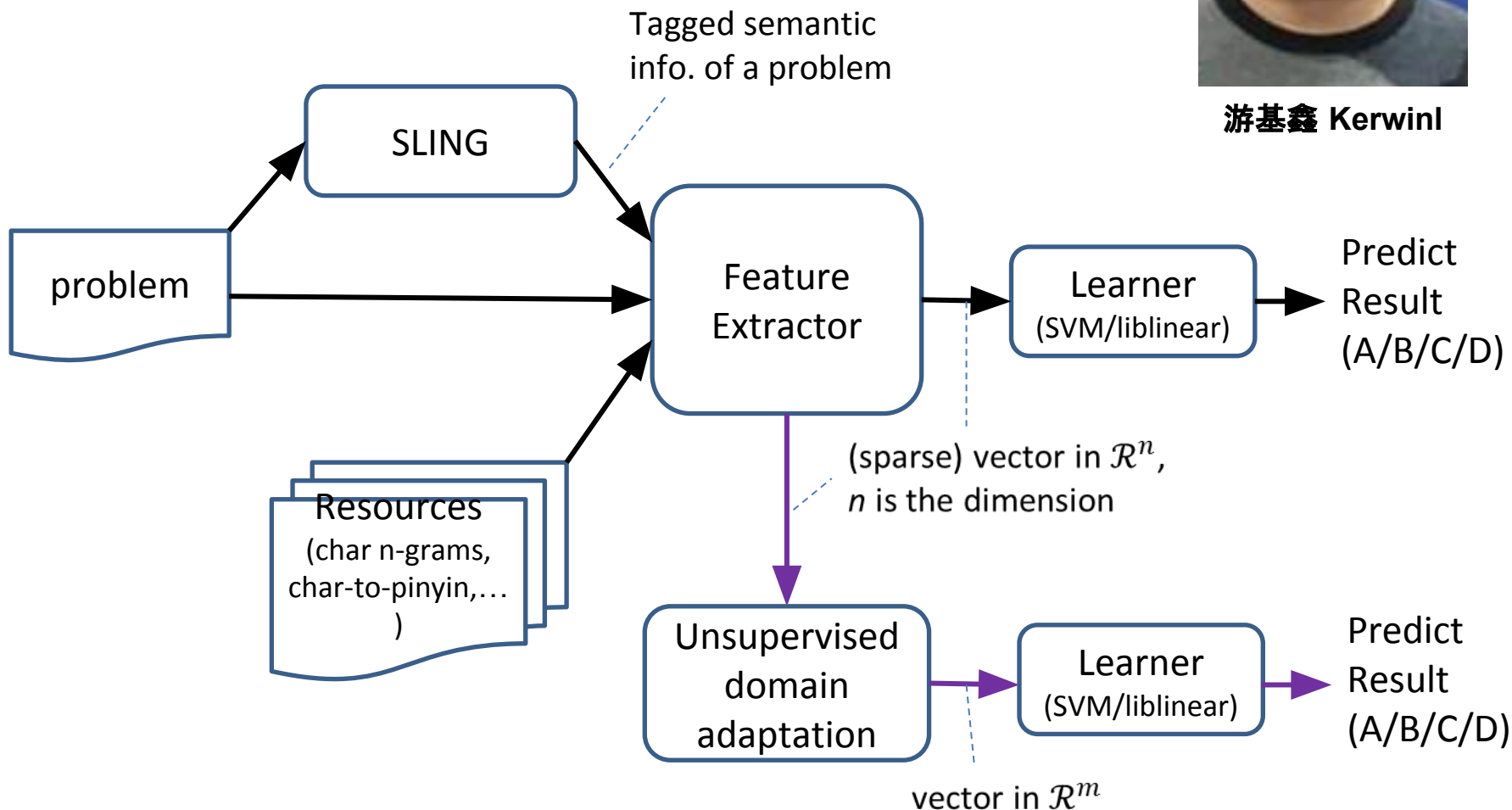
Paper of DiSAN:

<https://arxiv.org/abs/1709.04696>

SVM Sub-Learner



游基鑫 Kerwinl

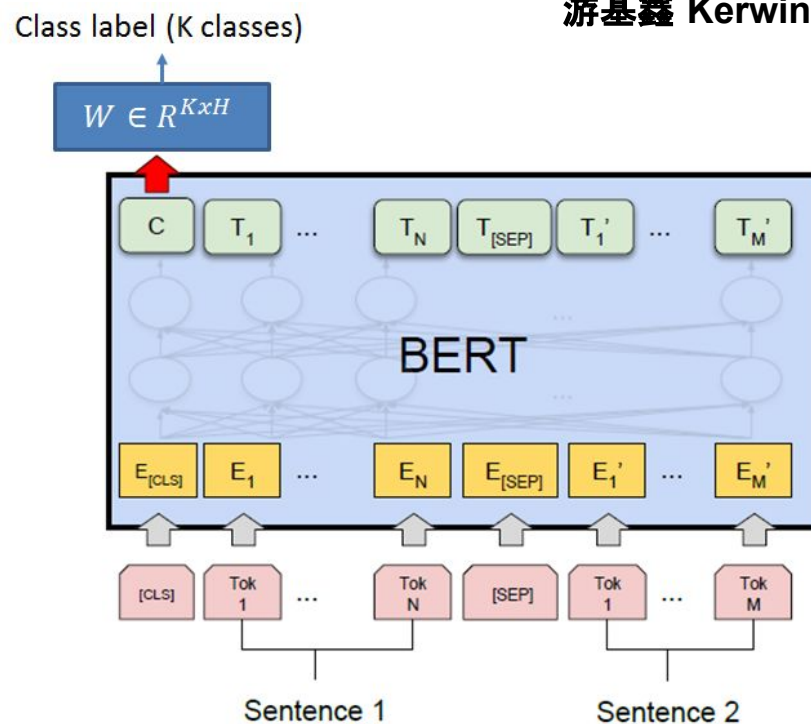
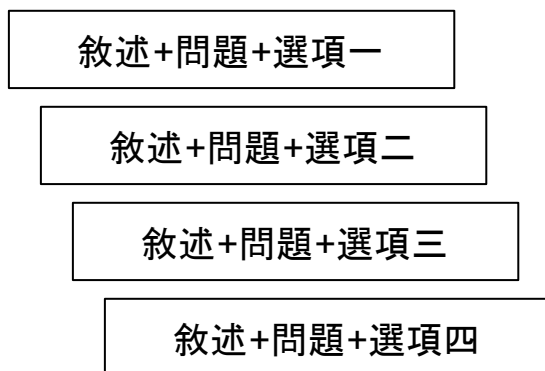


Google Bert

- Use Google BERT models



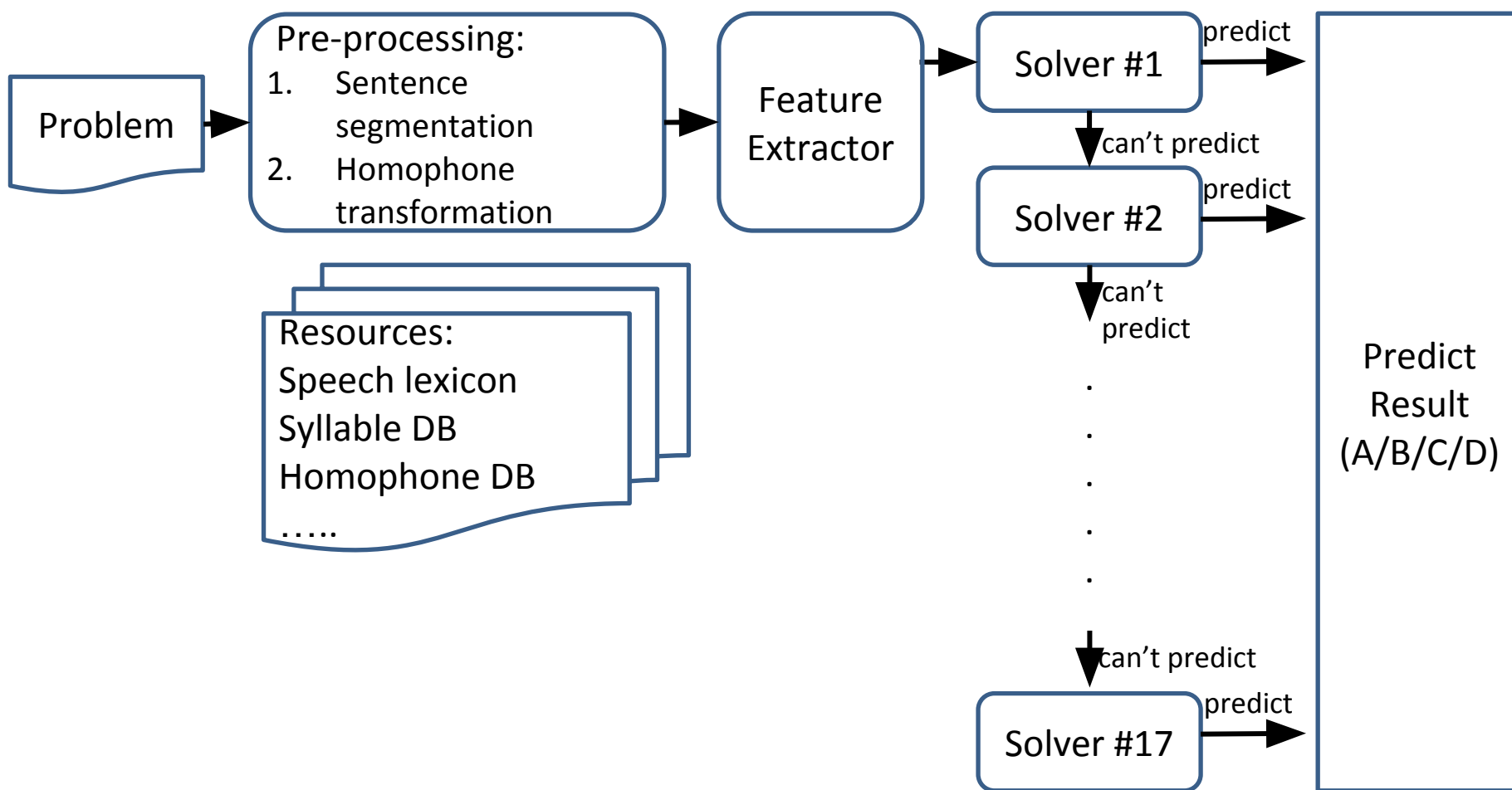
游基鑫 Kerwinl



(picture is modified from BERT, 2018, Figure 3)
<https://arxiv.org/abs/1810.04805>

Multi-Solver NLU

張瑋浩
Martin



Multi-Solver NLU

- 一時吃過了飯我就在繼之的工室桌上寫了一封回書交給帳房辭了繼之出來人到城裡去路上想著記我伯父的信已經有好幾天了不免去探問探問就順路走至我伯父公館先打聽回來。。。
- 若沒有說是還沒有回來我正要問我的信寄去了沒有忽然抬頭看見我那封信還是端端正正地插在一個壁架子上心中不覺暗暗動怒只不便同房門。。。
- 理論於是也不多言就走了回來細想這底下人。。。
- 可以這麼膽大應該寄的信。。。
- 也不拿上去回我伯母。。。
- 莫非繼之說的話。。。
- 當真不錯伯父有心避過了我嗎又想到就是伯父有心避過我這底下人也不該哥騎我的信難道我伯父交代過不可代我通信的嗎想來想去總想不出個道理。。。

Question:我寫了一封回書交給誰

1. 繼之
2. 伯父
3. 帳房
4. 伯母

Solution=15 - Guess=3 - Correct Ans=3

Implication Inference Engine



吳牧哲 Mick

- Believe-Desire-Intention (BDI) Model

• 連愛因斯坦都會解這個方程式。



• 連(小明/張三/我)都會解這個(方程式)。

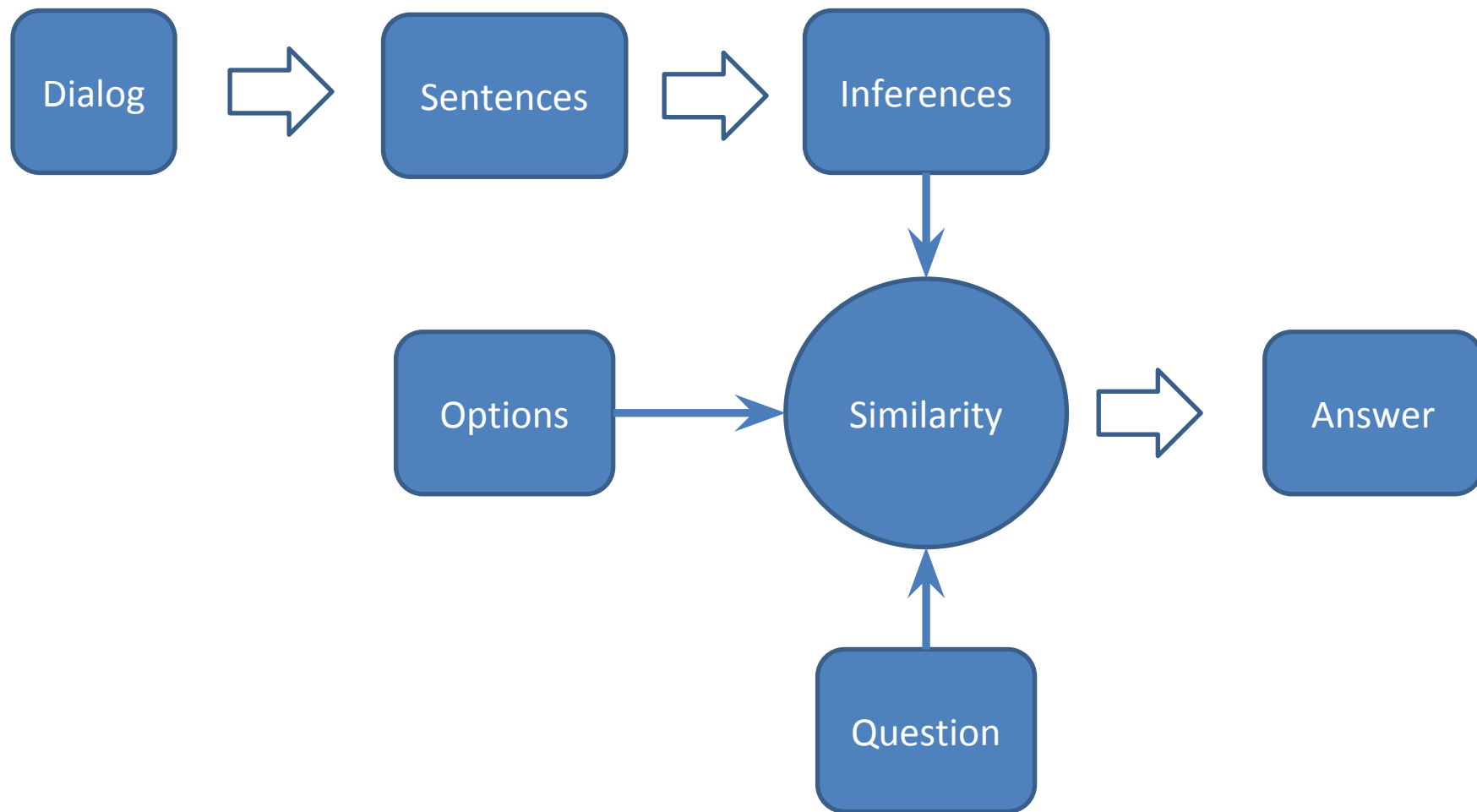


- 這個方程式並不難。
- 有愛因斯坦這個人。
- 有一個本句所提及的方程式。
- 愛因斯坦不善於解方程式。



- 這個(方程式)並不難。
- 有(小明/張三/我)這個人。
- 有一個本句所提及的(方程式)。
- (小明/張三/我)不善於解(方程式)。

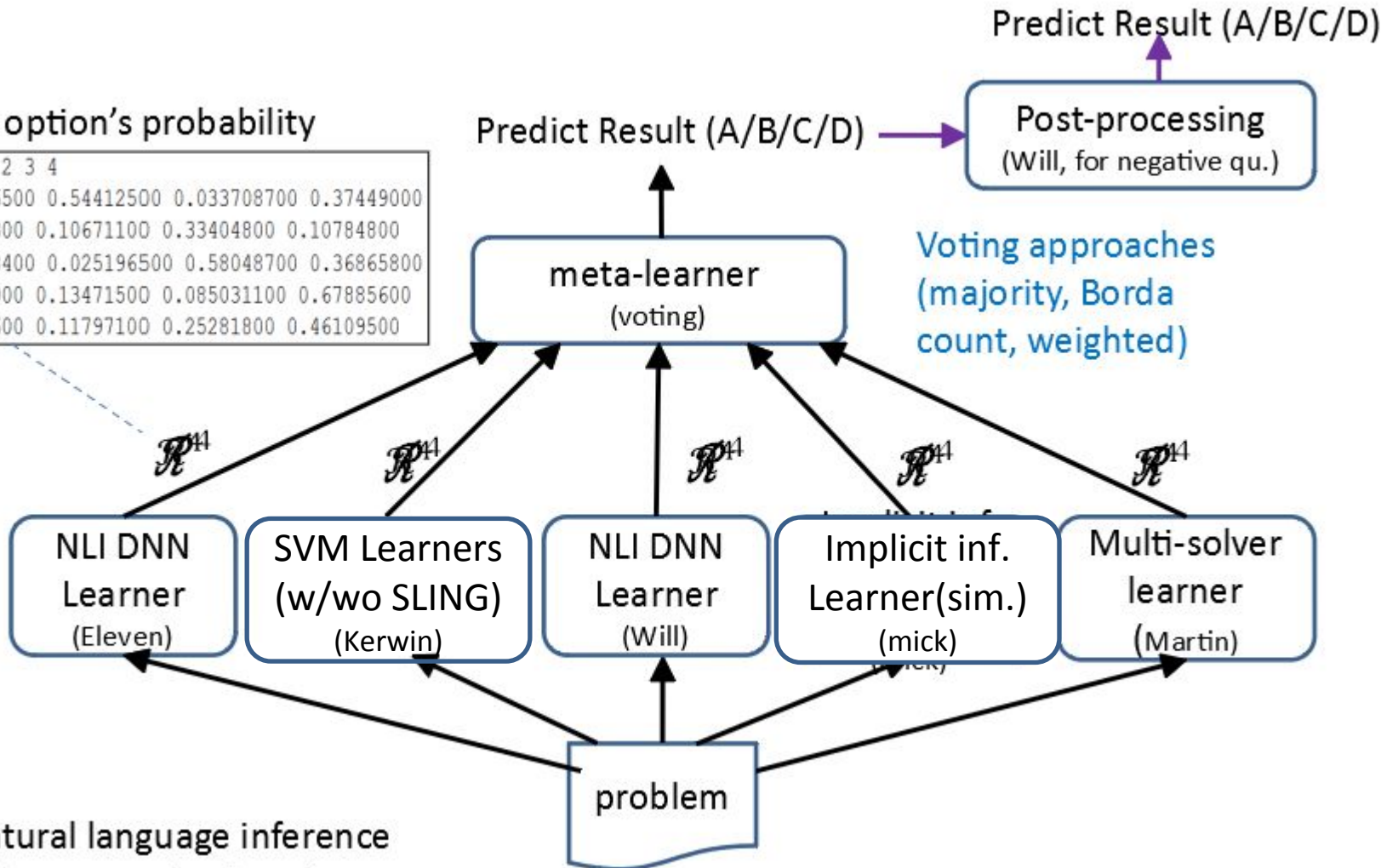
Implication Inference Engine



Emsemble

Format: option's probability

ID	answer	1	2	3	4
1	2	0.047676500	0.54412500	0.033708700	0.37449000
2	1	0.45139300	0.10671100	0.33404800	0.10784800
3	3	0.025658400	0.025196500	0.58048700	0.36865800
4	4	0.10139900	0.13471500	0.085031100	0.67885600
5	4	0.16811600	0.11797100	0.25281800	0.46109500



NLI: natural language inference
 DNN: deep neural network

Post-Processing

- Pattern Matching
 - 反向題
 - 以上皆是
 - 以上皆非

- 5% Improvement

簡答題

- 「因為...所以...」Pattern
- 人、時、地、物 Name Entity
- Summerization

Thanks

- Special Thanks:

華碩電腦 施崇棠董事長

華碩電腦 龐台銘技術長

- Speech Team:

Alim Misbullah (馬俊力)
Archer Tsai (蔡幸育)
BiCheng Yan (顏必成)
Eric Wu (吳安翔)
Husaini Muhammad (胡桑尼)
Jack Hsu (許時懷)
Ko-Cheng Wheng (翁恪誠)
Kyle Yu (游智凱)
Meng Wu (吳孟哲)
Pranav Jawale
Raymond Tsao (曹紘嘉)
Weiyuan Chen (陳威遠)
Seastar Lin (林海興)

- NLP Team:

Bill Chiu (邱元基)
Cap Huang (黃致凱)
CK Lin (林晉達)
Daniel Yeh (葉基善)
Eleven Liu (劉玠均)
Hudson Hsueh (薛家晟)
Jim Chung (鍾鴻源)
Kerwin Yu (游基鑫)
Martin Chang (張緯浩)
Mick Wu (吳牧哲)
Owen Luo (羅國峯)
Wei He (何浩璋)
Weichun Hu (胡維鈞)
Will Chen (陳黃威)
Sam Tu (涂耘昇)
Sigma Chen (陳立錚)

- Others:

Chaolun Chen (陳昭綸)
Jason Weng (翁億宗)
Mia Chiang (江瑋美)
Wayne Chung (鍾承恩)
Yitse Huang (黃毅澤)