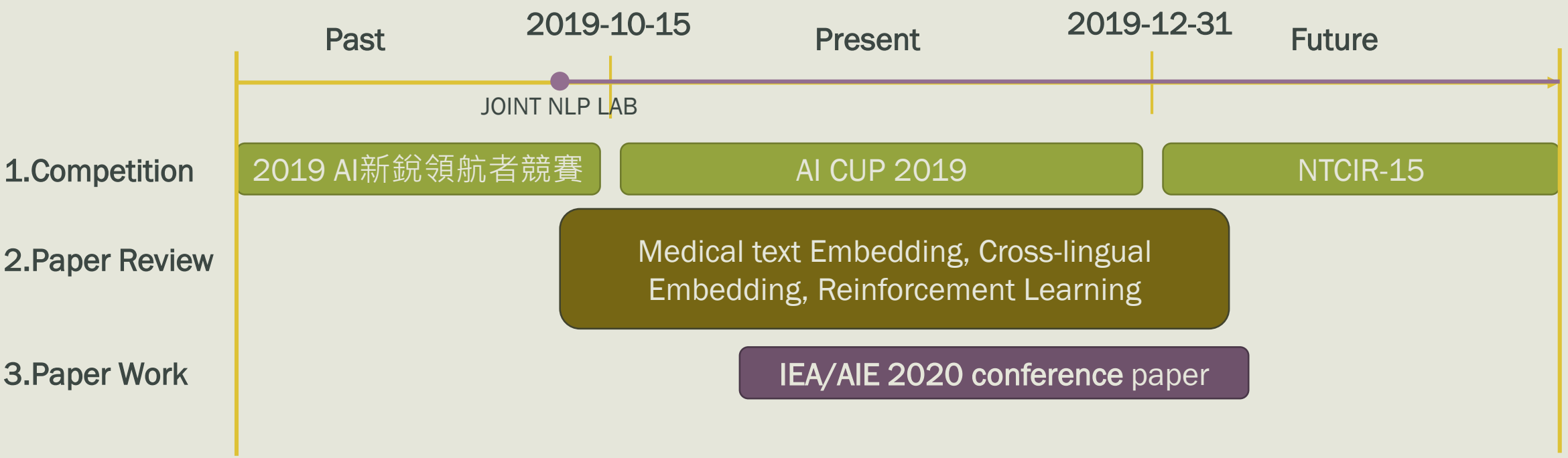
Junior students' presentation (12/18)

Student : 黃彥鈞 Weber YC, Huang (m946108006)
Director : Yung-Chun, Chang； Min-Huei, Hsu
Lab : TMU NLP

Content


Past
2019-10-15
Present
2019-12-31
Future
JOINT NLP LAB
1.Competition
2019 AI新銳領航者競賽
AI CUP 2019
NTCIR-15
2.Paper Review
Medical text Embedding, Cross-lingual Embedding, Reinforcement Learning
3.Paper Work
IEA/AIE 2020 conference paper

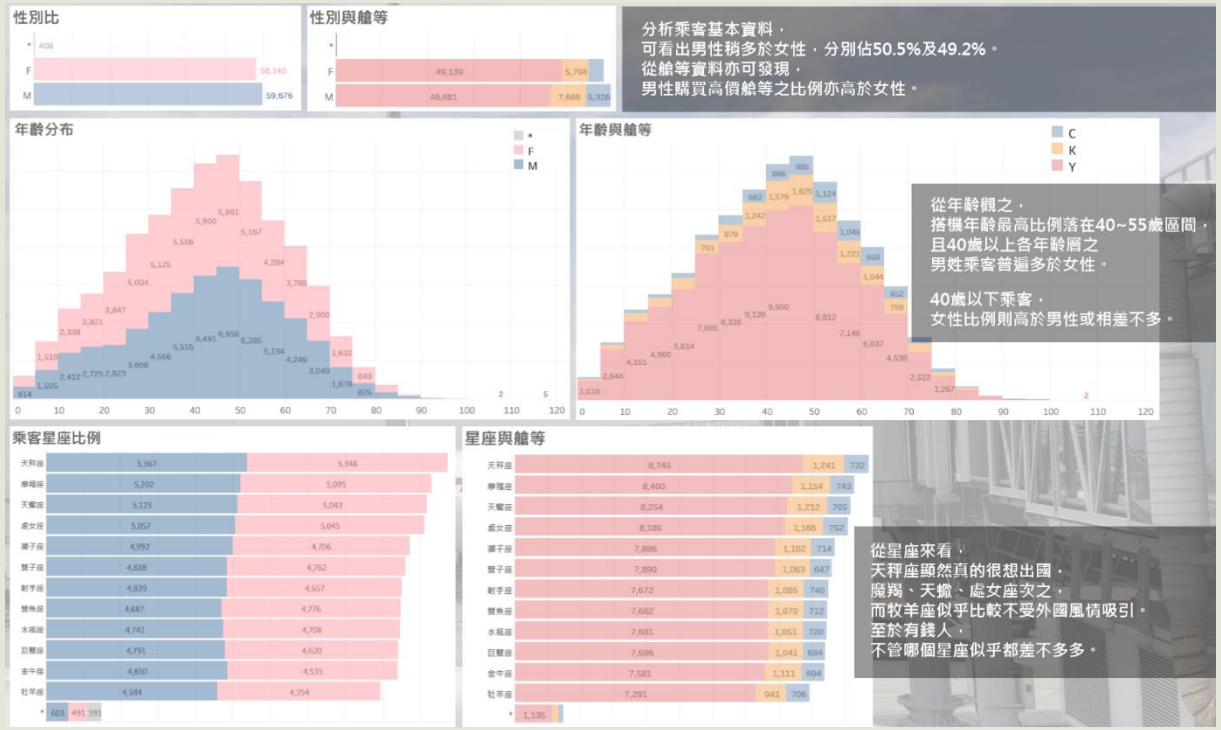
Competition

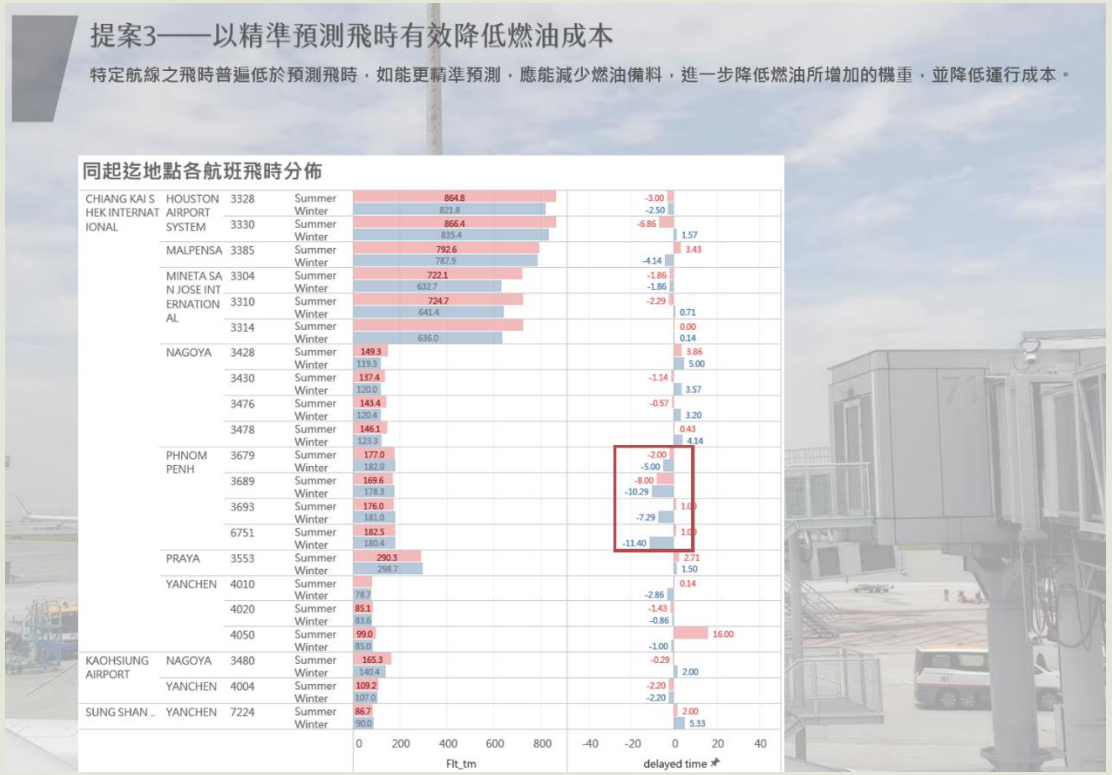
- 2019 AI 新銳領航者競賽
- AI CUP 2019
- NTCIR-15

2019 AI 新銳領航者競賽 – 長榮航空組 (1/2)

- 競賽目標 : 影響航空公司成本的原因很多，其中以燃油成本佔比最高，而燃油成本上升的 因素則包括班機總重、飛行時長.....等。請透過探索性資料分析報告找出相關因素，在飛行安全無虞及精緻服務品質的前提下，給予合適的建議以有效地控管該成本。
- Goal : There are many factors that affect flight costs, among which fuel costs account for the highest proportion, and factors that increase fuel costs include the total weight of flights, flight duration, etc. Given the ease and degradation of refined service quality, give appropriate advice to effectively control the cost.

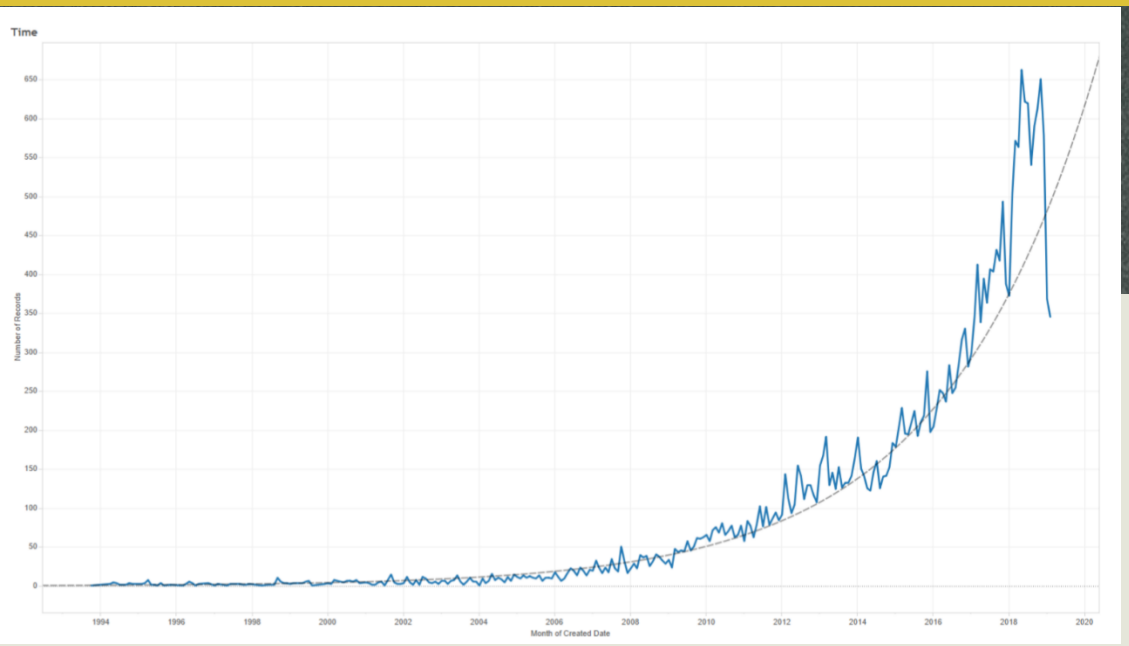
2019 AI 新銳領航者競賽 – 長榮航空組

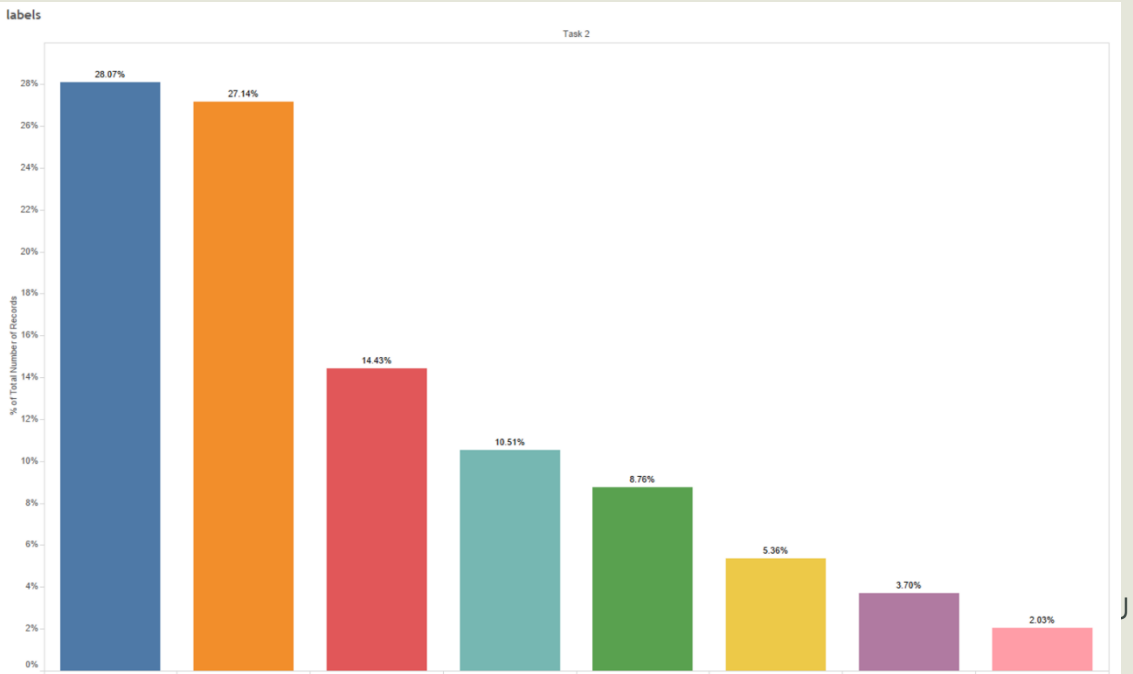


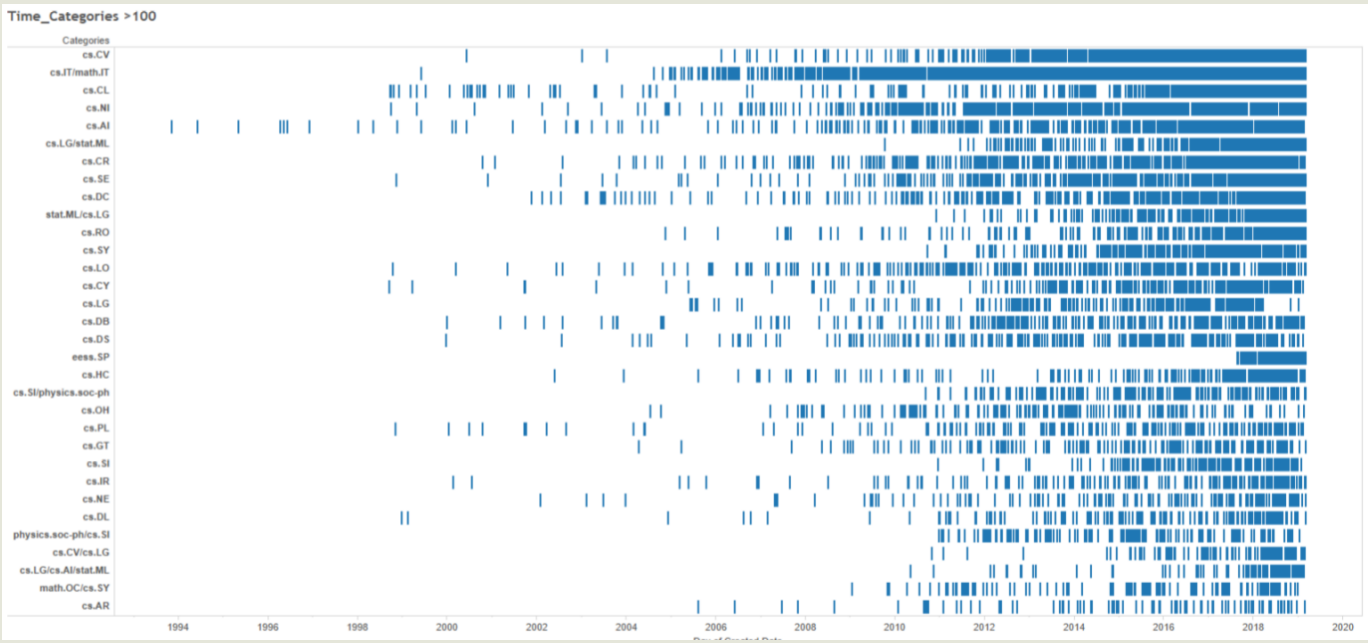


AI CUP 2019 (1/2)

- 教育部全國大專校院人工智慧競賽(AI CUP 2019)-人工智慧論文機器閱讀競賽之論文分類
- 競賽目標：從arXiv的電腦科學相關論文摘要，預測出摘要所屬的類別(Theoretical Paper, Engineering Paper, Empirical Paper, Others)。需注意的是摘要可以有多個分類，例如: 摘要可以同時是Theoretical Paper和Engineering Paper。
- Goal : From the abstract of arXiv computer science related paper, predicting the categories to which the abstract belongs (Theoretical Paper, Engineering Paper, Empirical Paper, Others). It should be noted that the abstract can have multiple categories, for example: The abstract can be both Theoretical Paper and Engineering Paper.


Time
Number of Records
Month of Created Date


labels
Task 2
% of Total Number of Records
28.07%
27.14%
14.43%
10.51%
8.76%
5.36%
3.70%
2.03%


Time_Categories >100
Categories
cs.CV
cs.IT/math.IT
cs.CL
cs.NI
cs.AI
cs.LG/stat.ML
cs.CR
cs.SE
cs.DC
stat.ML/cs.LG
cs.RO
cs.SY
cs.LO
cs.CY
cs.LG
cs.DB
cs.DS
eess.SP
cs.HC
cs.SI/physics.soc-ph
cs.OH
cs.PL
cs.GT
cs.SI
cs.IR
cs.NE
cs.DL
physics.soc-ph/cs.SI
cs.CV/cs.LG
cs.LG/cs.AI/stat.ML
math.OC/cs.SY
cs.AR

NTCIR-15 (NII Testbed Community for Information access Research)

- 競賽目標：

自1997年以來，NTCIR項目一直致力於增強信息訪問（IA）技術的研究工作，例如資訊檢索（IR），文本摘要，資訊提取（IE）和問題解答（QA）技術。其一般目的是：提供研究基礎架構，使研究人員可以對IA技術進行大規模評估；組成一個研究社區，在其中共享和交換基於可比實驗結果的發現；並制定IA技術的評估方法和性能度量。

- NTCIR-15特別著重於IA技術的兩個主題。
  - 資訊檢索：傳統及其他（DialEval，WWW-3，數據搜索和MART）
  - 從文本中獲取知識（FinNum-2，QA Lab-PoliInfo-2和SHINRA2020-ML）

- Goal :

Since 1997, the NTCIR project has promoted research efforts for enhancing Information Access (IA) technologies such as Information Retrieval (IR), Text Summarization, Information Extraction (IE), and Question Answering (QA) techniques. Its general purposes are to: Offer research infrastructure that allows researchers to conduct a large-scale evaluation of IA technologies；Form a research community in which findings based on comparable experimental results are shared and exchanged；Develop evaluation methodologies and performance measures of IA technologies.

- In particular, the NTCIR-15 focuses on two topics on IA technology mainly;
  - IR: Traditional and beyond (DialEval, WWW-3, Data Search, and MART)
  - Harvesting knowledge from text (FinNum-2, QA Lab-PoliInfo-2, and SHINRA2020-ML)

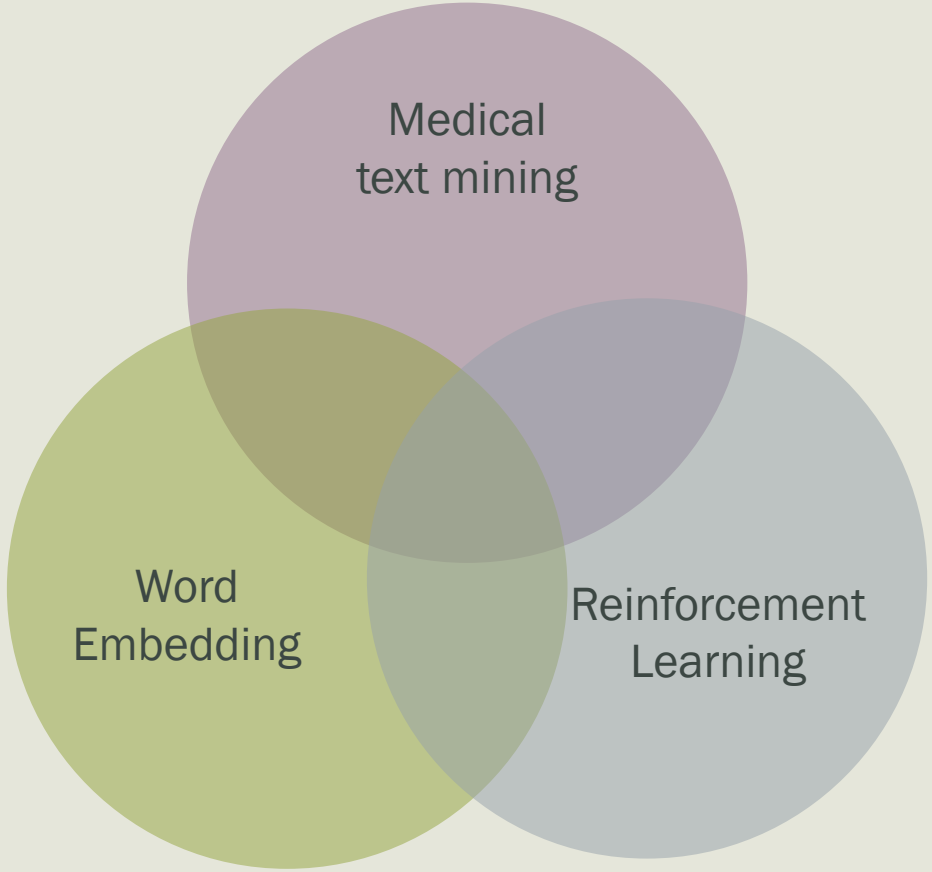
Paper Review

- Medical text Embedding
- Cross-lingual Embedding
- Reinforcement Learning

Previous Studies (1/2)

- Yang, Y., Zhang, Y., Tar, C., & Baldridge, J. (2019). PAWS-X: A Cross-lingual Adversarial Dataset for Paraphrase Identification. *arXiv preprint arXiv:1908.11828*.
- Goss, F. R., Plasek, J. M., Lau, J. J., Seger, D. L., Chang, F. Y., & Zhou, L. (2014). An evaluation of a natural language processing tool for identifying and encoding allergy information in emergency department clinical notes. In *AMIA Annual Symposium Proceedings* (Vol. 2014, p. 580). American Medical Informatics Association.
- Wang, H., Liu, X., Tao, Y., Ye, W., Jin, Q., Cohen, W. W., & Xing, E. P. (2019). Automatic Human-like Mining and Constructing Reliable Genetic Association Database with Deep Reinforcement Learning. In *PSB* (pp. 112-123).

Previous Studies (2/2)



Paper Work

- IEA/AIE 2020 conference paper

IEA/AIE 2020 conference paper

- IEA/AIE 2020 continues the tradition of emphasizing applications of applied intelligent systems to solve real-life problems in all areas including engineering, science, industry, automation & robotics, business & finance, medicine and biomedicine, bioinformatics, cyberspace, and human-machine interactions.

IEA/AIE 2020 conference paper

Introduction

- Social media topic(text-mining)
- Sentiment analysis

Related work surveying

- Different method applied in sentiment and opinion analysis
- Word Embedding developments

Method building

- TFIDF + 6ML
- WE + CNN/RNN
- BERT-base(fine-tuning)
- BERT-base + CNN
- LLR-BERT (main focus)

Experiment

- ENG : Kaggle movie review dataset
- CHN : Reader emotion dataset
- CHN : PTT movie review dataset

Thanks for listening !!! Q & A ?