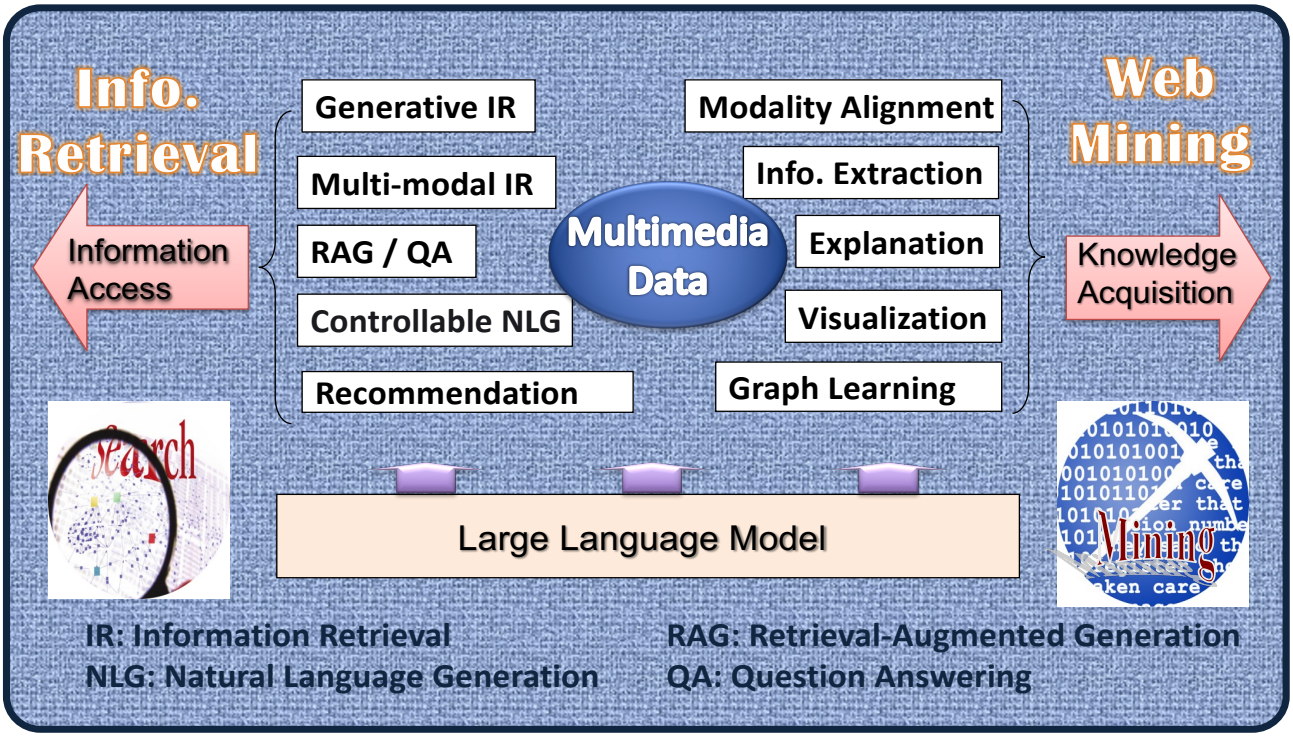
網路探勘與資訊檢索實驗室
Web Mining & Information Retrieval Lab

PI: Pu-Jen Cheng 鄭卜壬
網頁: http://www.csie.ntu.edu.tw/~pjcheng
信箱: pjcheng@csie.ntu.edu.tw

歡迎有興趣的同學加入
請準備履歷等基本資料
寄至左側信箱或至R220



實驗室簡介:

網路探勘與資訊檢索實驗室是一個充滿創新與合作精神的研究環境，致力於探索網路資料的潛力以及資訊檢索的最新技術。「網路探勘」探討如何從大規模網路資料中發掘有價值的知識，而「資訊檢索」研究如何預測使用者需求與資料間的相關性，從而提升搜尋引擎的準確性與效率，這兩個領域相輔相成，透過應用深度學習模型、大型語言模型，共同推動資訊技術的前沿發展。

研究層面涵蓋語意理解、語意搜尋、資訊擷取、推薦系統、文本生成、意見分析與生成、社群網路分析、使用者行為分析與預測、多模態資料與自然語言處理、知識圖譜建構與推理、反事實推理與因果分析等理論及應用(醫療、行銷、法律等)。

近期主要研究方向:

- Generative Information Retrieval (生成式資料檢索模型)
- Generative Recommender Systems (生成式推薦系統)
- Multi-modal / Cross-modal IR & RecSys (多模態/跨模態之資訊檢索與推薦系統)
- Controllable Natural Language / Text Generation (可控制的自然語言/文本生成)
- Enhancing Explainability in Modeling (強化深度學習模型的可解釋性)
- Retrieval-Augmented Generation (檢索增強生成)
- Semantic Understanding & Generation of Special Data Structures (Graph, Table,…) (特殊資料結構(如圖型、表格)的語意理解與生成)

近期部份研究主題

Product Recommendations Using Multi-Modal Reviews

(CIKM 2023) (ICTIR 2023)

Review Comments $\mathcal{R}$

Interacted Items $\mathcal{I}$ with visual and textual contents $\mathcal{T}, \mathcal{V}$

Users $\mathcal{U}$

... I felt my energy levels rise and my thirst was quenched. I highly recommend ....

Gatorade Thirst Quencher Sports Drink, Variety Pack, 20oz Bottles, 12 Pack, Electrolytes for Rehydration

... The cheese was super salty. Not buying again.

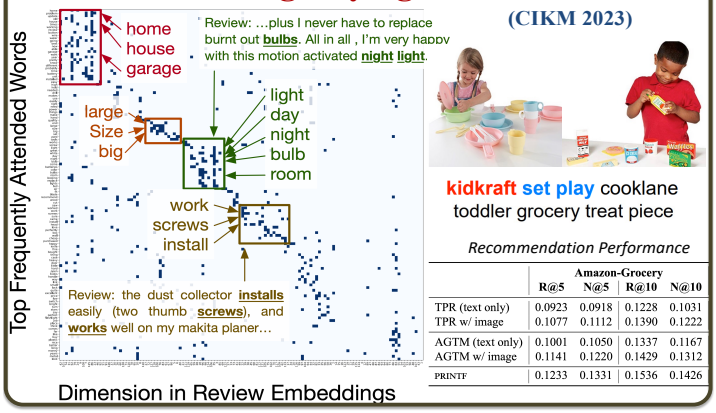
RITZ Handi-Snacks Crackers and Cheese Dip, 6 - 0.95 Ounce Packs (Pack of 6)

... For the price per unit paying under a dollar a pack at time of writing these are pretty good.

Wonderful Pistachios, No Shells, Variety Pack, 0.75oz (Pack of 9) Roasted & Salted (4), Chili Roasted (3), Honey Roasted (2)

Discovering Buying Preferences

(CIKM 2023)



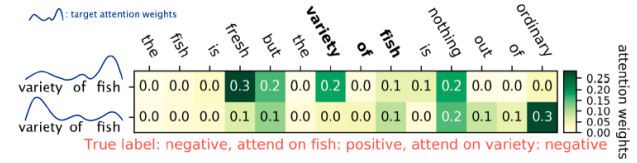
kidkraft set play cooklane toddler grocery treat piece

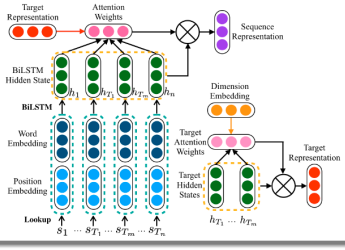
Recommendation Performance

	Amazon-Grocery			
	R@5	N@5	R@10	N@10
TPR (text only)	0.0923	0.0918	0.1228	0.1031
TPR w/ image	0.1077	0.1112	0.1390	0.1222
AGTM (text only)	0.1001	0.1050	0.1337	0.1167
AGTM w/ image	0.1141	0.1220	0.1429	0.1312
PRINTF	0.1233	0.1331	0.1536	0.1426

Multi-aspect Sentiment Analysis

(RecSys 2020)



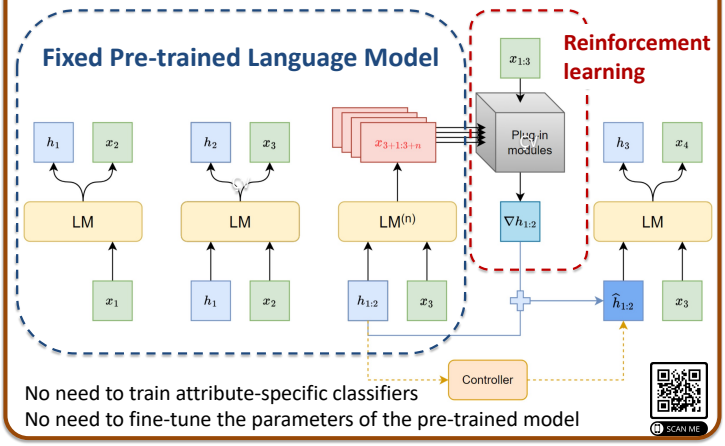


Exploring Different Aspects

謝謝護理人員用心照顧我，在住院期間用無限的寬容幫助我，也謝謝身心科主任不斷醫治我與住院醫師。不過缺點是地下室停車場很難停車。

Controlling Text Generation

(NAACL Findings 2024)

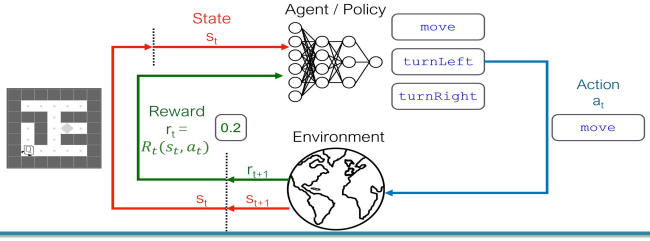


Generating Negative Comments for Sentiment Control

Once upon a time….

Method	Outputs
GPT-2 (OpenAI, 2019)	Once upon a time, I lived in the city. I loved it. I was so happy there. It felt so peaceful. I would go there every day to work, go to the movies, go to concerts. I'd go to the park. There was nothing
PPLM (ICLR 2020)	Once upon a time, a while back I., a woman had a pregnancy and a man was in a lot of people, and it's not a lot of the same, the best, I think that's my brother has the best of it, there's a
FUDGE (NAACL 2021)	Once upon a time no (or else... or worse, i might even let bad-favoriti (i.a/i/a/i/a/something/anya /i.a / or /
PREADD (ACL Findings 2023)	Once upon a time revelation Starg exempt aspects inravings believearl capable prosecuted fromDownload 109thoroughlyeki NonethelessDallas towering animalslevant +— Watkins Fah protrisome solicement guess immune satisfactoryuggestabilityxml WitatechacetPokÖ0e9monmidt cooper stocks plenty known 433studos realization bored dwind
PiLM-RL	Once upon a time, a person who was so depressed that he couldn't even function properly, and couldn't find a place in this world for himself, and couldn't even get any help from anyone, couldn't take any of it anymore, and couldn't even get
PiLM-Controller	Once upon a time, I could not believe that so many men could be so evil. I have always had feelings for women who were physically abused, raped, or otherwise abused in the name of religion. But the world is so much worse than what we imagine.

Explainability in Reinforcement Learning

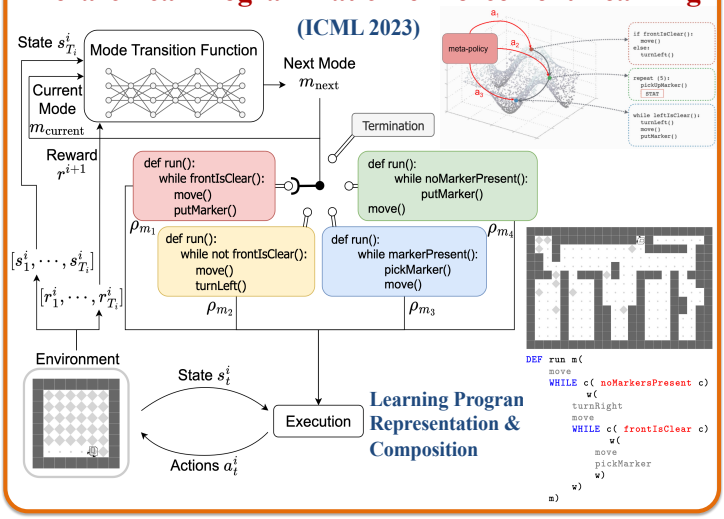


醫院評價

(SIGIR 2019)

評價：+3, +2, +1, -1, -2, -3，空白指未提及。

評價	醫術	服務	環境	設備	交通	飲食	整體
繳費可以用機器、App 很方便。冷氣不太冷，門診區有點悶。			-1	+2			+1
住院品質極差。病患不能好好休息。空調真的很吵，工作人員吵鬧聊天大叫，真的無言了。 2018/05(新) 至護理站反應隔壁床衛生習慣不好，大便丟廁所又關冷氣，熱到滿頭汗，請他規勸，護理士說要我們自己處理並且又說病患要惡搞也不關他們的事，叫我自己寫建議書。路過的一位好心護士幫忙跟隔壁規勸也改善了。○○也是有認真幫忙病患的醫護人員，但領薪水不做事的也是有。		-2	-3				-2
新院區明亮舒服,服務親切,小兒科和耳鼻喉科醫師都很細心,特別是掛○院長過敏免疫科,雖然等候多時,但是醫術高明,找到主因,治療舒緩孩子過敏問題,很感恩! 小兒科的醫師也都很親切溫柔。急診的醫護人員也很用心。 107 年 7 月因家人闌尾炎開刀自費一天 2500 入住核醫大樓兩人房病房,訂膳食,覺得住院已經好辛苦,吃到膳食更痛苦了,湯品完全是洗碗水,髮菜像髮絲在裡面,早餐兩天吃到麵跟炒米粉,完全覺得膳食很不 ok 啊！如果醫院人員也三餐吃膳食應該就能了解。	+3	+3	+3			-2	+1
設備齊全，醫師技術也很好。 不過偶爾會被實習人員當練習台，技術不到位一再重試，感覺並不好。	-1			+3			-2
婦產科○○○醫師 非常專業 讓病人很放心 ○醫師細心又親切把病人當成自己家人一樣。 住院期間 ○○病房的護士 也很盡責照顧病患 不過白天真的不太好停車，搭火車 .公車 .計程車 或許較合適	+3	+3			-1		+2

Hierarchical Programmatic Reinforcement Learning

(ICML 2023)