Natural Language Processing and Reasoning

Strong AI Lab & LIU AI Lab, School of Computer Science,

The University of Auckland

Presented by

Qiming Bao

My Ph.D. Supervisors: Professor Michael Witbrock and Dr. Jiamou Liu

30th August 2022

Strong AI Lab & LIU AI Lab

- Strong AI Lab is led by Professor Michael Witbrock, at the intersection of machine learning, reasoning, and natural language understanding, with an additional focus on achieving the best social and civilisational impacts of increasingly powerful AI.
- LIU AI Lab is led by Dr. Jiamou Liu. We are an AI research group at the University of Auckland. We are engaged in artificial intelligence research and development from both the industrial and the academic sides. Our research interests cover a wide range of topics across the modern AI world, including deep learning, reinforcement learning, multi-agent systems, natural language processing, and complex network analysis.

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

https://www.ai.ac.nz/sail/
https://www.liuailab.org/

What have we achieved?

- Strong AI Lab has published more than 24 top-rank conferences and journals, including IJCAI, ACL, AAAI/EAAI, COLING, NAACL, AAMAS, ICML, PAKDD, IJCLR, AJCAI and so on. [1] won the best presentation award for AJCAI.
- LIU AI Lab has published more than 47 top-rank conferences and journals, including AAMAS, SIGIR, UAI, NIPS, ICML, IJCAI, AAAI/EAAI, NAACL, ICDM, EUSIPCO, PAKDD, IJCLR, KSS and so on. Some of them have high citations, like Mandy [2] has more than 88 citations.
- We have applied for several international and national grants and strong international/national research cooperation, and we welcome any students(Honours/MSc/Ph.D.) or postdocs to join our labs!

THE UNIVERSITY OF AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND

https://www.ai.ac.nz/sail/
https://www.liuailab.org/

[1] Yogarajan, V. 2022. Predicting COVID-19 Patient Shielding: A Comprehensive Study. AJCAI.
[2] L Ni et al. 2017. Mandy: Towards a smart primary care chatbot application. International symposium on knowledge and systems sciences.

Natural Language Processing (NLP)

- Natural language processing is an important direction in artificial intelligence. It studies various theories and methods enabling effective communication using natural language between humans and computers.
- Natural language processing involves natural language, that is, the language that people use every day, so it is closely related to the study of linguistics, but it is very different. Natural language processing is not the general study of natural language but the study of computer systems, especially software systems and algorithms, that can effectively realize natural language communication.
- Natural language processing has many application scenarios, including machine translation, automatic summarization, text classification, question answering, text semantic comparison, etc.

The Development of NLP

2018 to now - Transformer and Pretrained Langauge Model [5]

↑

2008 to 2017 - Word2Vec, RNNs, LSTMs, GRUs for Deep Learning [3] [4]

↑

1970s to early 2000s - Statistics-based Approach [2]

↑

1950s to 1970s - Rules-based Approach [1]

[1] Warren Weaver. 1955. Translation. Machine Translation of Languages, 14:15-23, 1955.
[2] Peter F. Brown et al. 1990. A Statistical Approach to Machine Translation. Comput. Linguistics, 16, 79-85.
[3] Chung, J. et al. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, NIPS 2014.
[4] Mikolov, T. et al. 2013. Efficient estimation of word representations in vector space.
[5] Qiu, X. et al. 2020. Pre-trained Models for Natural Language Processing: A Survey, Science China Technological Sciences, 63(10), 1872-1897.

Rules-based Approach

- The earliest research work in natural language processing was machine translation. In 1949, Weaver first proposed the design of machine translation [1].
- The methods at that time were limited to computation resources, fixed rules, and linguistic and syntax grammar, while the variation of natural language was very diverse, and only limited problems were handled.

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

[1] Warren Weaver. 1955. Translation. Machine Translation of Languages, 14:15-23, 1955.

Statistics-based Approach

- Frederick Jelinek et al. [1] from IBM proposed a statistic-based method to avoid design complex rules and templates to process language in machine translation.

$$P(S) = P(w_1, w_2, \cdots, w_n)$$

$$P(w_1, w_2, \ldots, w_n) = P(w_1)P(w_2 \mid w_1)P(w_3 \mid w_1, w_2) \ldots P(w_n \mid w_1, w_2, \ldots, w_{n-1})$$

Example:

S1: I am happy today.
S2: I today am happy.
S3: Happy I am today.

S1 is more likely to be chosen as a correct sentence than S2 and S3.

[1] Peter F. Brown et al. 1990. A Statistical Approach to Machine Translation. Comput. Linguistics, 16, 79-85.

Statistics-based Approach

- Markov proposed an effective assumption that the probability of occurrence of any word w_i is only related to its preceding word w_{i-1}
- The bigram statistic language Model is based on this assumption. We can extend that to the n-gram model, which will cost more computation resources.

$$P(S) = P(w_1) \cdot P(w_2 \mid w_1) \cdot P(w_3 \mid w_2) \cdots P(w_i \mid w_{i-1}) \cdots P(w_n \mid w_{n-1})$$

Unigram: when n=1
Bigram: when n=2
Trigram: when n=3

This is Big Data AI Book

Uni-Gram	This	Is	Big	Data	AI	Book
Bi-Gram	This is	Is Big	Big Data	Data AI	AI Book	
Tri-Gram	This is Big	Is Big Data	Big Data AI	Data AI Book		

[1] Peter F. Brown et al. 1990. A Statistical Approach to Machine Translation. Comput. Linguistics, 16, 79-85.
[2] https://www.oreilly.com/library/view/artificial-intelligence-for/9781788472173/c11d01e5-bf13-488f-8922-16f6028254e4.xhtml

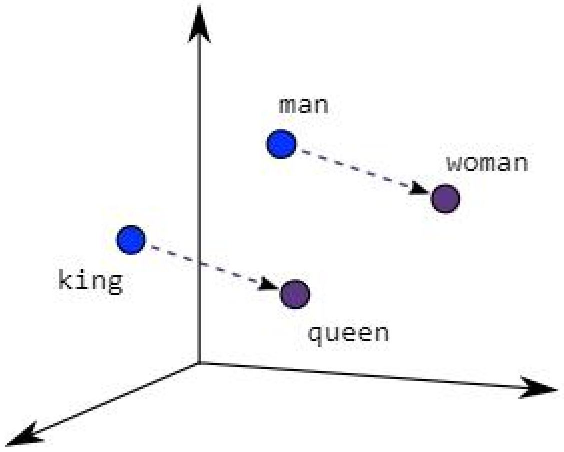
Word Embedding

- How to transfer word into vector representation?
  - One-hot encoding
    - One word, one encoding (vectors are too sparse and vectors are orthogonal.)
    - The dimension of the vector (dimension explosion)

word	**1**	**0**	**0**	**0**	**0**	**0**
words	**0**	**1**	**0**	**0**	**0**	**0**

Word2Vec

- How to transfer word into vector representation?
  - Word2Vec
    - CBOW: Use context to predict the center word.
    - Skip-gram: Use the center word to predict context.

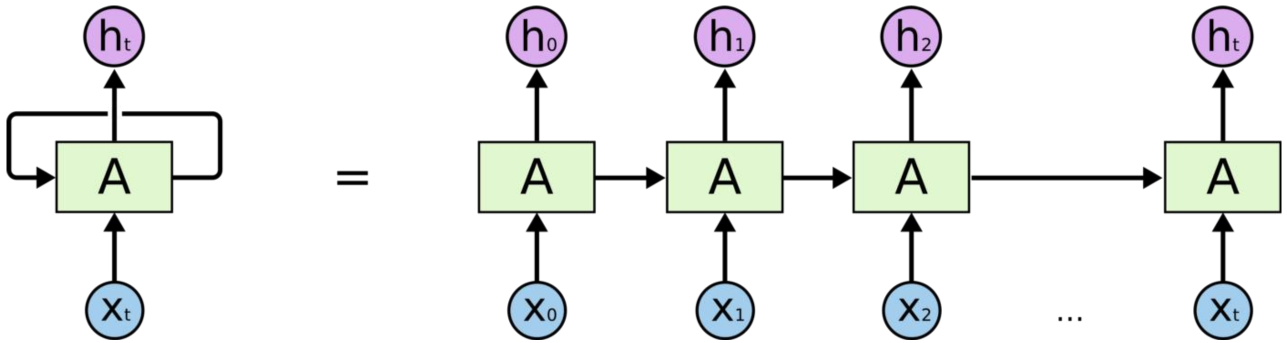


Words with similar meanings have closer distances [1].

[1] https://www.analyticsvidhya.com/blog/2021/07/word2vec-for-word-embeddings-a-beginners-guide/

Recurrent Neural Network (RNN)

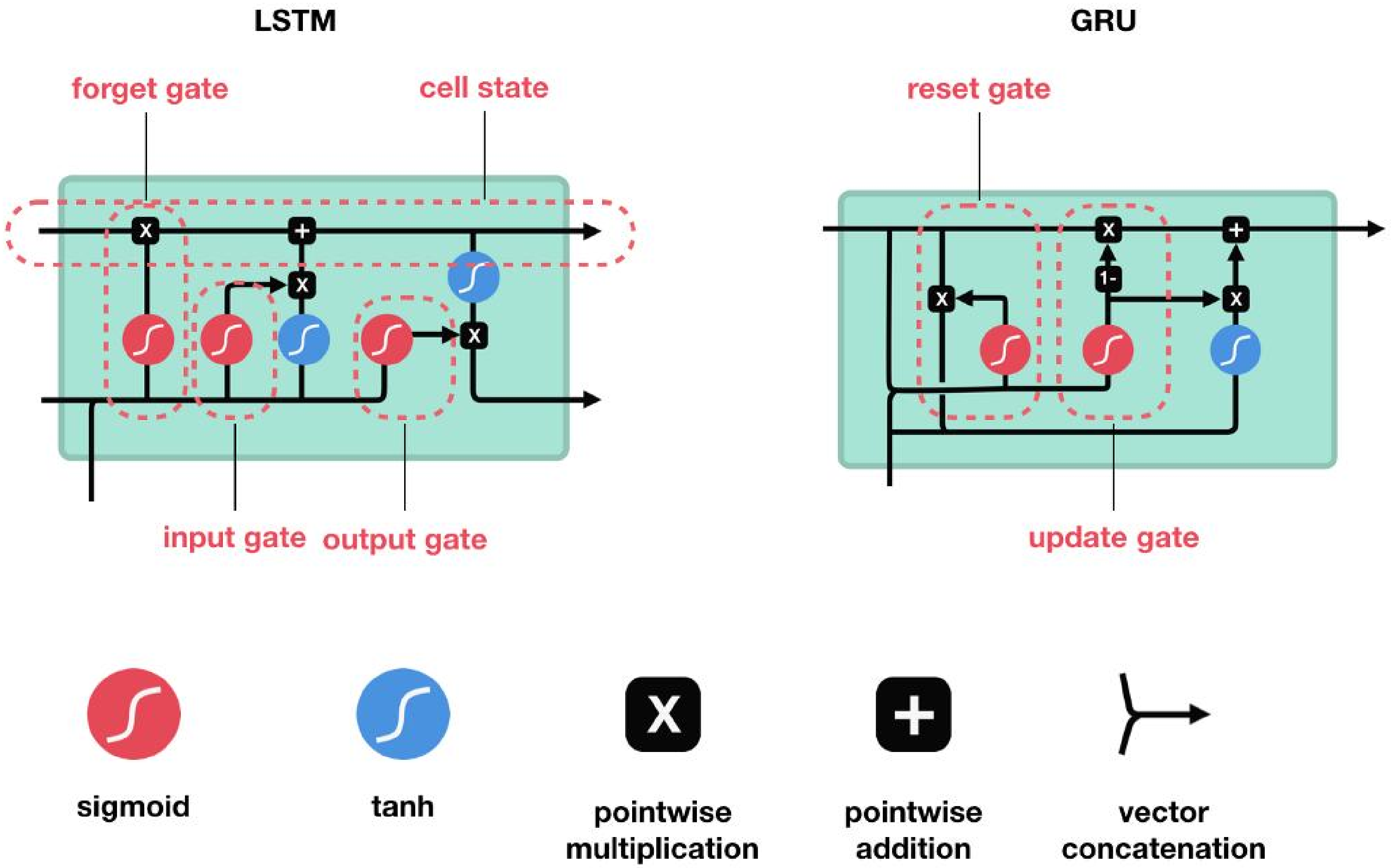
- RNN is proposed to better address the problem of sequence learning and sequence translation, that is, the problem of word-to-word sequence.
- For example, given sentence “high-speed trains drive at high speed on the rails.”, swapping the order of each word in this sentence changes the meaning of the sentence.



J J Hopfield et al. Neural networks and physical systems with emergent collective computational abilities. PNAS 1982.
https://towardsdatascience.com/introduction-to-recurrent-neural-network-27202c3945f3

Long/Short Term Memory (LSTM) & GRU

- When the sentence length is very long, RNN will suffer from gradient vanishing and gradient exploding. To solve these problems, there are two existing methods, including LSTM and GRU.



THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

Hochreiter, S. et al. 1997. LONG SHORT-TERM MEMORY. Neural computation, 9(8), 1735-1780.
Cho, K. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. EMNLP.
https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

Transformer

- Transformer was proposed in [1] which is a multi-head self-attention architecture. Transformer adopts global dependencies, which breaks through the limitation that RNN-based models cannot be parallelized, and can better handle long-distance text tasks.
- A large number of pre-trained language models including BERT [2], GPT-2 [3], GPT-3 [4], etc., adopt the Transformer architecture, and achieve state-of-the-art results in machine translation, reading comprehension, question answering and other tasks.

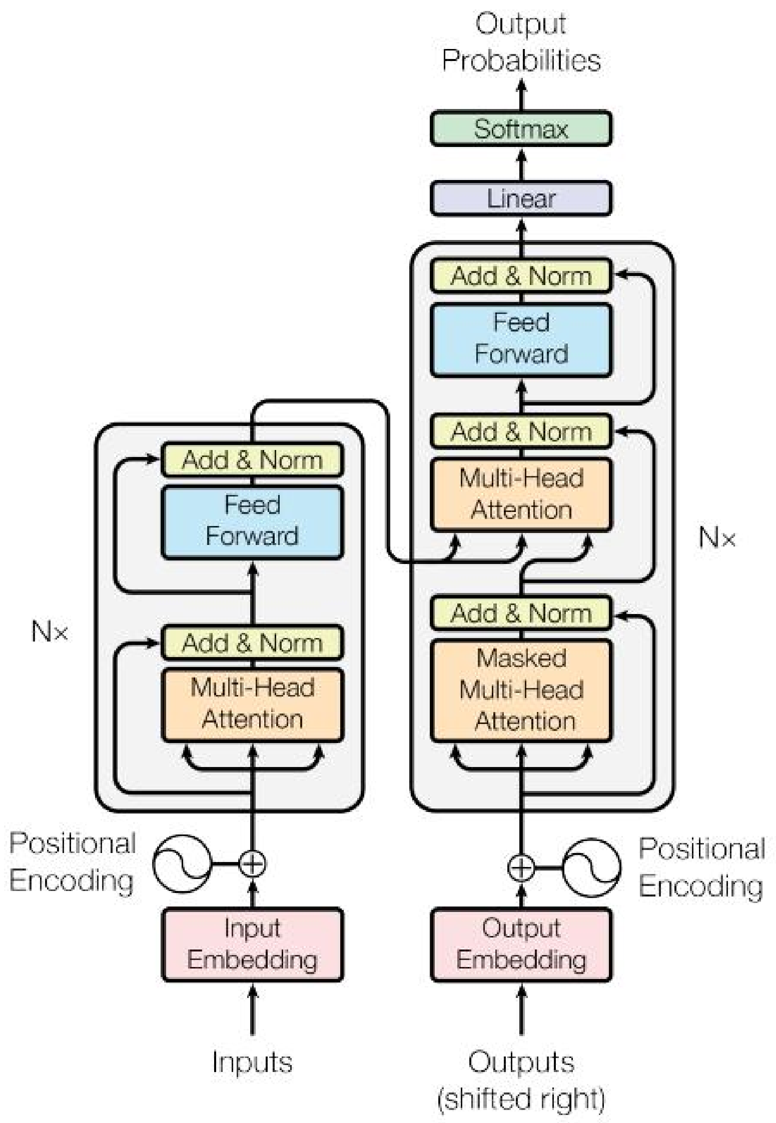


Figure 1: The Transformer - model architecture.

[1] Vaswani, A. et al. 2017. Attention is all you need. NIPS.
[2] Devlin, J. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL 2018.
[3] Radford, A. 2019. Language Models are Unsupervised Multitask Learners. OpenAI blog.
[4] Brown, T. et al. 2020. Language Models are Few-Shot Learners. NIPS 2020.

Symbolic Logic Reasoning (SLR)

- ***Symbolic logic*** is a set of formal languages that deals with how symbols express meaning and how reasoning can be carried out through symbolic means.
- Symbolic logic example:
  - Propositions:
    - (A) Bob is big.
    - (B) All big people are rough.
    - (C) Bob is rough.
  - Conclusion:
    - $A \land B \Rightarrow C$
  - The $\land$ means “and,” and the $\Rightarrow$ symbol means “implies.”

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

Symbolic Logic Programs

- A normal logic program expressed by Prolog [1]

$$p(X) :- q(X).$$

$$q(a).$$

$p(X)$, where variables are upper case characters

$q(a)$, where ground truth are lower case.

[1] Programming in Prolog: Using the ISO standard, Clocksin, 2012

Symbolic Logic Programs

1: Facts

$e(l).$

$?e(l).\ 1$

$?i(d).\ 0$

2: Unification

$o(V,V).$

$?o(d,d).\ 1$

$?o(b,d).\ 0$

3: 1 Step

$p(X):-q(X).$

$q(a).$

$?p(a).1$

$?p(b).0$

[1] Cingillioglu, N. et al., 2018. DeepLogic: Towards End-to-End Differentiable Logical Reasoning, AAAI-MAKE19.

Natural Language Reasoning

- We can understand natural language reasoning as representing symbolic logic in natural language.
  - **Logic of natural language**, such as propositional logic and first-order logic.
  - **Diversity and flexibility of natural language**, such as polysemy, a paraphrase of sentences.
  - Reasoning requires understanding existing information to obtain unknown information.

    Deductive reasoning: Given premise and rules to derive the conclusion.

    Inductive reasoning: Given premise and conclusion to derive rules.

    Abductive reasoning: Given rules and conclusion to derive premise.

    More examples can be found in [1] and [2].

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

[1] Bao, Q. et al., 2022. Multi-Step Deductive Reasoning Over Natural Language: An Empirical Study on Out-of-Distribution Generalisation. IJCLR-NeSy.
[2] Young, N. et al. 2022. AbductionRules: Training Transformers to Explain Unexpected Inputs. The finding of ACL.

Example for Natural Language Reasoning

(Input Facts:) Alan is blue. Alan is rough. Alan is young.
Bob is big. Bob is round.
Charlie is big. Charlie is blue. Charlie is green.
Dave is green. Dave is rough.

(Input Rules:) Big people are rough.
If someone is young and round then they are kind.
If someone is round and big then they are blue.
All rough people are green.

Q1: Bob is green. True/false? **[Answer: T]**
Q2: Bob is kind. True/false? **[F]**
Q3: Dave is blue. True/false? **[F]**

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

[1] Clark, P., et al. 2020. Transformers as Soft Reasoners over Language, IJCAI 2020.

Example for Natural Language Reasoning

(Input Facts:) Alan is blue. Alan is rough. Alan is young.
Bob is big. Bob is round.
Charlie is big. Charlie is blue. Charlie is green.
Dave is green. Dave is rough.

(Input Rules:) Big people are rough.
If someone is young and round then they are kind.
If someone is round and big then they are blue.
All rough people are green.

Q1: Bob is green. True/false? **[Answer: T]**
Q2: Bob is kind. True/false? **[F]**
Q3: Dave is blue. True/false? **[F]**

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

[1] Clark, P., et al. 2020. Transformers as Soft Reasoners over Language, IJCAI 2020.

Example for Natural Language Reasoning

(Input Facts:) Alan is blue. Alan is rough. Alan is young.
Bob is big. Bob is round.
Charlie is big. Charlie is blue. Charlie is green.
Dave is green. Dave is rough.

(Input Rules:) Big people are rough.
If someone is young and round then they are kind.
If someone is round and big then they are blue.
All rough people are green.

Q1: Bob is green. True/false? **[Answer: T]**
Q2: Bob is kind. True/false? **[F]**
Q3: Dave is blue. True/false? **[F]**

THE UNIVERSITY OF AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND

[1] Clark, P., et al. 2020. Transformers as Soft Reasoners over Language, IJCAI 2020.

Example for Natural Language Reasoning

(Input Facts:) Alan is blue. Alan is rough. Alan is young.
Bob is big. Bob is round.
Charlie is big. Charlie is blue. Charlie is green.
Dave is green. Dave is rough.

(Input Rules:) Big people are rough.
If someone is young and round then they are kind.
If someone is round and big then they are blue.
All rough people are green.

Q1: Bob is green. True/false? **[Answer: T]**
Q2: Bob is kind. True/false? **[F]**
Q3: Dave is blue. True/false? **[F]**

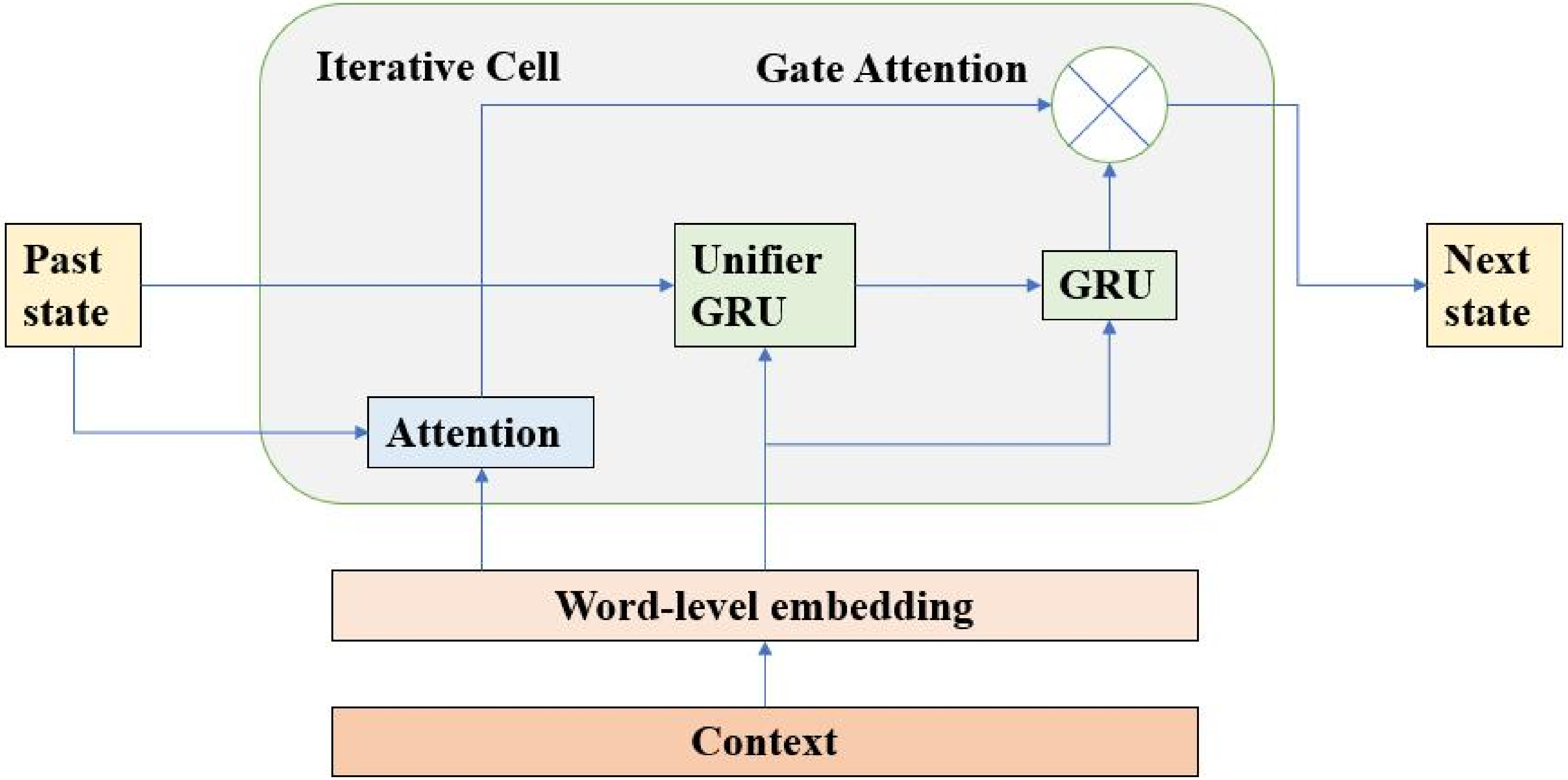
[1] Clark, P., et al. 2020. Transformers as Soft Reasoners over Language, IJCAI 2020.

Research Gap

- Existing models, including DeepLogic and other RNN-based baseline models, have room for improvement in their reasoning abilities over natural language.
- We found existing models are not good at out-of-distribution (OOD) generalisation, including three aspects:
  - When model be trained on shallow reasoning depth set and tested on the deeper reasoning depth.
  - When model be trained on synthetic natural langauge dataset and tested on the dataset with extra paraphrased natural language rewritten by human.
  - When model be trained on unshuffled dataset and tested on shuffled dataset.
- Existing multi-step deductive reasoning datasets like PARARULES and CONCEPTRULE V1 and V2 have unbalanced distributions on the shallow and deep reasoning depth. The deeper depth (2<=Depth<=5) has much less dataset than the shallow depth (Depth <2).

[1] Bao, Q. et al., 2022. Multi-Step Deductive Reasoning Over Natural Language: An Empirical Study on Out-of-Distribution Generalisation. IJCLR-NeSy.

Model Overview



[1] Bao, Q. et al., 2022. Multi-Step Deductive Reasoning Over Natural Language: An Empirical Study on Out-of-Distribution Generalisation. IJCLR-NeSy.

Word-level Embedding

- The word vector representation is based on GloVe [1].
- The input representation layer to the network is a sequence of two word-level sentences used for context and statement, respectively.
- The concatenation of context and statement representation will be feed into the gated recurrent unit (GRU).

THE UNIVERSITY OF AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND

[1] Pennington, et al., 2014. Glove: Global vectors for word representation, EMNLP.

Iteration

- The iteration process is from the DeepLogic [1]. The iteration step consists of attending to the rules, computing a new state using each rule and **updating** the old state.
- To apply a rule, we use another recurrent neural network called the **inner GRU unifier** that processes every literal of every rule. The inner GRU unifier needs to learn unification between **variables** and **constants** as well as how each rule interacts with the current state.

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

[1] Cingillioglu, N. et al., 2018. DeepLogic: Towards End-to-End Differentiable Logical Reasoning, AAAI-MAKE19.

Gate Attention

- Dynamic Memory Network+ [1] achieved 100% test accuracy by using gate attention on bAbI deductive reasoning task (Task-15), which gave us the idea of integrating Gate Attention into DeepLogic. GRU can use gate attention to update the internal state.

[1] Xiong, C., et al., 2016. Dynamic Memory Networks for Visual and Textual Question Answering, ICML.

Establish Baselines - RNNs & PLM

- We have three baseline models that we borrowed from the bAbI task leaderboard. We also set DeepLogic as one of the baseline methods, and then we have a Transformer-based model RoBERTa-Large as a baseline model. We use glove.6B.zip [4] as the word vector representation for the RNN-based models.
  - Long short-term memory (LSTM, 1997) [1] (The baseline method on bAbI dataset),
  - Dynamic Memory Network (DMN, 2016) [2] (One of the first paper use Attention in the memory network),
  - Memory Attention Control networks (MAC, 2018) [3] (A classical method from memory network).

[1] Hochreiter, et al. 1997. Long short-term memory,
[2] Kumar, et al. 2016. Ask me anything: Dynamic memory networks for natural language processing, ICML
[3] Hudson, et al. 2018. Compositional attention networks for machine reasoning, ICLR.
[4] Pennington, et al. 2014. Glove: Global vectors for word representation, EMNLP.
[5] Liu, Y. et al., 2019. Roberta: A robustly optimized bert pretraining approach. arxiv.

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

CONCEPTRULE vs CONCEPTRULE V2

(*Input Context:*) Book is not located in bed.
Bed is located in loft.
Loft is located in city.
City is located in fast-food restaurant.
Question 1: Book is located in loft. True/False? [**Answer: T**]
Question 2: Bed is located in city. True/False? [**Answer: T**]
Question 3: Book is located in bed. True/False? [**Answer: F**]

(*Input Context:*) Book is not located in bed.
Bed is located in loft.
Loft is located in city.
City is located in fast-food restaurant.
Question 1: Book is not located in bed. True/false? [**Answer: T**] [**Depth: 0**]
Question 2: Book is not located in loft. True/false? [**Answer: T**] [**Depth: 1**]
Question 3: Book is not located in city. True/false? [**Answer: T**] [**Depth: 2**]

THE UNIVERSITY OF AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND

Hartill, T., CONCEPTRULE, https://drive.google.com/file/d/1lxoAvtcvqVCYiO8e3tENnrTQ1NNVtpjs/view
Hartill, T., CONCEPTRULE V2, https://drive.google.com/file/d/1lOCbW8bfZxj1RIzKDxn8xKg99XyYNj7z/view

Dataset Description

Table 2

Information about the datasets used in this paper. PARARULES has less number of examples that require deep reasoning steps. CONCEPTRULES V2 does not consider reasoning depths greater than 3. The train, dev and test set are already splitted by the author of each dataset.

Dataset		Depth=0	Depth=1	Depth=2	Depth=3	Depth=4	Depth=5
PARARULES	Train	290435	157440	75131	48010	9443	7325
	Dev	41559	22276	10833	6959	1334	1038
	Test	83119	45067	21496	13741	2691	2086
PARARULE-Plus	Train	-	-	89952	90016	90010	90022
	Dev	-	-	16204	16154	16150	16150
	Test	-	-	2708	2694	2704	2692
CONCEPTRULES V2 (full)	Train	2074360	1310622	873748	436874	-	-
	Dev	115148	72810	48540	24270	-	-
	Test	115468	72810	48540	24270	-	-
CONCEPTRULES V2 (simplified)	Train	131646	74136	49424	24712	-	-
	Dev	7166	4116	2744	1372	-	-
	Test	7362	4116	2744	1372	-	-

[1] Bao, Q. et al., 2022. Multi-Step Deductive Reasoning Over Natural Language: An Empirical Study on Out-of-Distribution Generalisation. IJCLR-NeSy.

Dataset Description

Table 3
The entity types and relation types for CONCEPTRULES V1 (simplified/full), CONCEPTRULES V2 (simplified/full), PARARULES, and our PARARULE-Plus.

Dataset	#Entity	#Relation	Shuffled Rules	Depth Tag	Derivable	NAF
CONCEPTRULES V1 (simplified)	385	7	No	No	Yes	Yes
CONCEPTRULES V1 (full)	4048	24	Yes	No	Yes	No
CONCEPTRULES V2 (simplified)	385	7	No	Yes	Yes	Yes
CONCEPTRULES V2 (full)	4048	24	Yes	Yes	Yes	Yes
PARARULES	19	4	No	Yes	Yes	Yes
PARARULE-Plus	71	8	No	Yes	Yes	Yes

THE UNIVERSITY OF AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND

[1] Bao, Q. et al., 2022. Multi-Step Deductive Reasoning Over Natural Language: An Empirical Study on Out-of-Distribution Generalisation. IJCLR-NeSy.

A Sample for Negation as Failure (NAF)

(*Input Facts:*) The bear visits the lion.
The tiger likes the cat.
The cat does not like the bear.
The lion likes the tiger.
(*Input Rules:*) If someone sees the lion then the lion is kind.
If the tiger visits the lion and someone does not see the tiger then the tiger visits the bear.
If someone likes the bear and they like the tiger then the bear visits the tiger.
If someone is not round then they like the cat.
If someone visits the lion then they are blue.
If someone visits the bear and they do not see the lion then they visit the tiger.
If someone is cold and they do not visit the lion then the lion visits the tiger.
If someone visits the tiger and they are green then the tiger likes the cat.
Question 1: The bear likes the cat. True/false? [**Answer: T**]
Question 2: The bear is round. True/false? [**F**]
Question 3: The bear is not round. True/false? [**T**]

THE UNIVERSITY OF AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND

[1] Peter Clark, et al. 2020. Transformers as Soft Reasoners over Language, IJCAI 2020.

Experiment Result

Table 4
We use GloVe [16] as the word vector representation. We use PARARULES with all depths as the training set for all models and then test them on examples with different reasoning depths (D). Comparison among our IMA-GloVe-GA, IMA-GloVe, MAC-GloVe, DMN-GloVe, IMASM-GloVe, LSTM-GloVe, and RoBERTa-Large on PARARULES test sets with different reasoning depths.

Train ↓; Test →	D=1	D=2	D=3	D≤3	D≤3+NatLang	D≤5	D≤5+NatLang
IMA-GloVe	0.861	0.853	0.830	0.842	0.810	0.792	0.705
MAC-GloVe	0.792	0.776	0.750	0.763	0.737	0.701	0.652
DMN-GloVe	0.846	0.843	0.817	0.827	0.789	0.779	0.666
IMASM-GloVe	0.864	0.855	0.824	0.838	0.801	0.782	0.608
LSTM-GloVe	0.500	0.500	0.500	0.499	0.499	0.500	0.500
IMA-GloVe-GA	**0.950**	**0.943**	**0.919**	**0.927**	**0.883**	**0.879**	**0.741**
RoBERTa-Large	**0.986**	**0.985**	**0.977**	**0.979**	**0.972**	**0.967**	**0.949**

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

[1] Bao, Q. et al., 2022. Multi-Step Deductive Reasoning Over Natural Language: An Empirical Study on Out-of-Distribution Generalisation. IJCLR-NeSy.

Experiment Result

Table 5

IMA-GloVe, IMA-GloVe-GA, and RoBERTa-Large trained on CONCEPTRULES V1 (simplified / full) and tested on different test sets. Rules in CONCEPTRULES V1 Simplified are not shuffled, while CONCEPTRULES V1 full contains randomly shuffled rules. CONCEPTRULES V1 full has larger number of relations and entities than CONCEPTRULES V1 simplified.

Model	Train set	Test accuracy (Simplified Test set)	Test accuracy (Full Test set)
IMA-GloVe	Simplified	0.994	0.729
	Full	0.844	**0.997**
IMA-GloVe-GA	Simplified	**0.998**	**0.747**
	Full	0.851	**0.999**
RoBERTa-Large	Simplified	**0.997**	0.503
	Full	**0.927**	0.995

THE UNIVERSITY OF AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND

[1] Bao, Q. et al., 2022. Multi-Step Deductive Reasoning Over Natural Language: An Empirical Study on Out-of-Distribution Generalisation. IJCLR-NeSy.

Experiment Result

Table 6

IMA-GloVe, IMA-GloVe-GA, and RoBERTa-Large trained on CONCEPTRULES V2 (full) and tested on test sets that require different depths of reasoning.

Model	Test set	Mod1 Depth=1	Mod2 Depth=2	Mod3 Depth=3	Mod01 Depth≤1	Mod012 Depth≤2	Mod0123 Depth≤3
IMA-GloVe	Depth=1	**0.999**	**0.998**	**0.990**	**0.997**	**0.998**	**0.997**
	Depth=2	**0.998**	**0.999**	**0.988**	**0.995**	**0.998**	**0.997**
	Depth=3	**0.997**	0.998	0.981	**0.991**	0.996	**0.997**
IMA-GloVe-GA	Depth=1	0.993	0.996	0.987	0.987	0.991	**0.997**
	Depth=2	0.993	**0.999**	0.974	0.986	0.991	0.995
	Depth=3	0.988	**1**	**0.994**	0.989	**0.997**	0.994
RoBERTa-Large	Depth=1	0.998	0.975	0.831	0.995	0.975	0.971
	Depth=2	0.997	0.972	0.885	0.993	0.972	0.965
	Depth=3	0.987	0.951	0.984	0.988	0.951	0.936

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

[1] Bao, Q. et al., 2022. Multi-Step Deductive Reasoning Over Natural Language: An Empirical Study on Out-of-Distribution Generalisation. IJCLR-NeSy.

Experiment Result

Table 7

RoBERTa-Large trained on PARARULES with different reasoning depths and tested on test sets that require different depths of reasoning. A bold number indicates the highest accuracy in a test set.

Model	Test set	Mod012 (Depth≤2)	Mod0123 (Depth≤3)	Mod0123Nat (Depth≤3+NatLang)	Mod012345 (Depth≤5)
RoBERTa-Large	Depth=0	**0.971**	0.946	0.968	0.953
	Depth=1	**0.943**	0.907	0.933	0.909
	Depth=2	**0.933**	0.902	0.932	0.902
	Depth=3	0.562	0.902	**0.926**	0.907
	Depth=4	0.481	0.863	**0.904**	0.888
	Depth=5	0.452	0.856	0.916	**0.933**
	NatLang	0.573	0.579	**0.962**	0.594

THE UNIVERSITY OF AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND

[1] Bao, Q. et al., 2022. Multi-Step Deductive Reasoning Over Natural Language: An Empirical Study on Out-of-Distribution Generalisation. IJCLR-NeSy.

Experiment Result

Table 8

RoBERTa-Large is fine-tuned on examples with different depths from PARARULES and also the entire PARARULE-Plus(PPT), and then is evaluated on test sets that require different depths of reasoning. The yellow background indicates improvement on accuracy after adding our PARARULE-Plus in the training process.

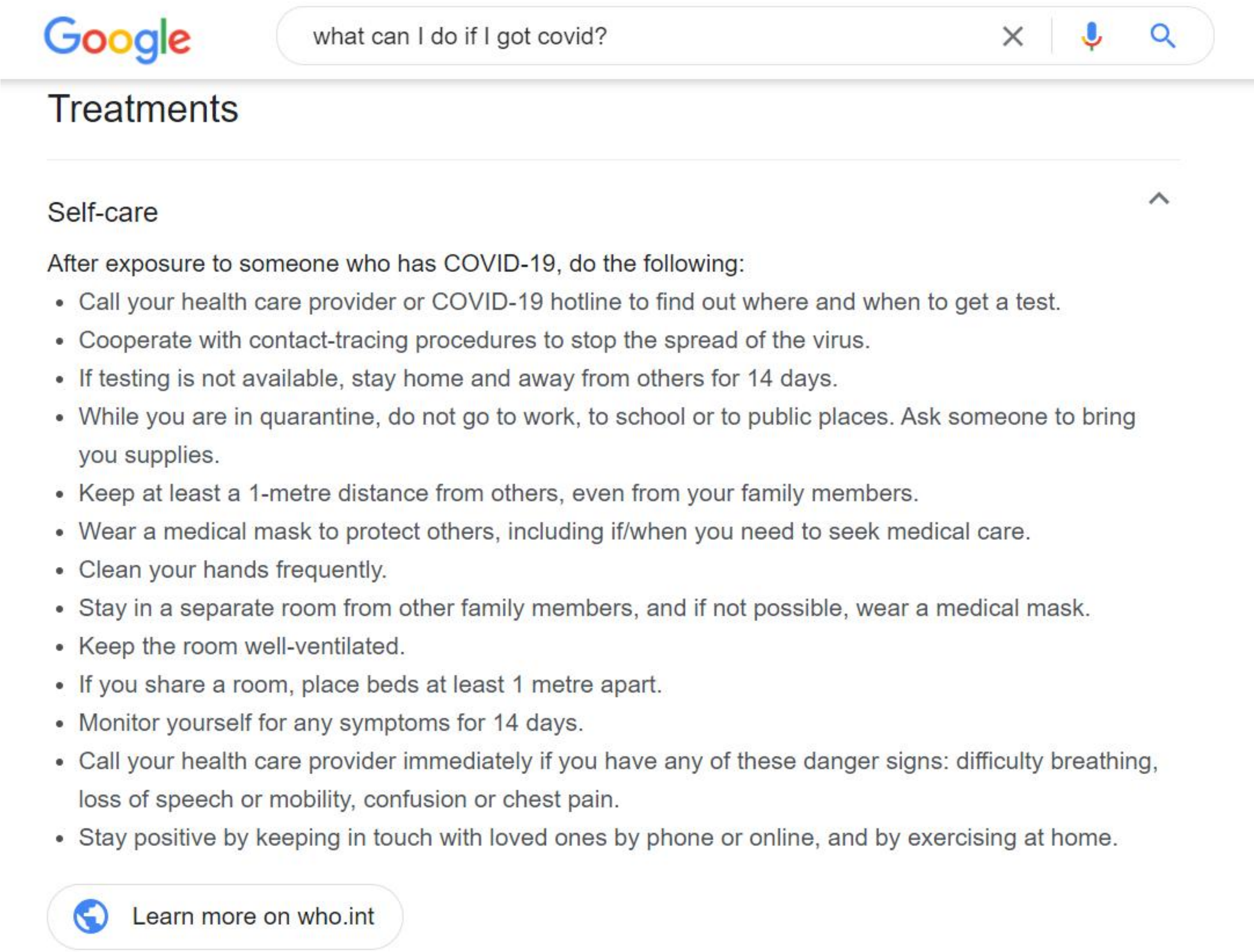
Model	Test set	Mod012 (Depth≤2+PPT)	Mod0123 (Depth≤3+PPT)	Mod0123Nat (Depth≤3+NatLang+PPT)	Mod012345 (Depth≤5+PPT)
RoBERTa-Large	Depth=0	0.946	0.901	0.965	**0.963 (+0.010)**
	Depth=1	0.877	0.847	**0.937 (+0.004)**	0.881
	Depth=2	0.868	0.873	**0.927**	0.839
	Depth=3	0.771 (+0.209)	0.862	**0.904**	0.826
	Depth=4	0.675 (+0.194)	0.852	**0.897**	0.832
	Depth=5	0.661 (+0.209)	0.888 (+0.032)	0.923 (+0.007)	**0.934 (+0.001)**
	NatLang	0.557	0.593 (+0.014)	**0.970 (+0.008)**	0.649 (+0.055)

THE UNIVERSITY OF AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND

[1] Bao, Q. et al., 2022. Multi-Step Deductive Reasoning Over Natural Language: An Empirical Study on Out-of-Distribution Generalisation. IJCLR-NeSy.

Application Scenarios

- Search engine like Google, Bing, Baidu. Google has announced that they use BERT to improve the search quality [1].



[1] https://blog.google/products/search/search-language-understanding-bert/

Machine Translation

- Google Translation [1] and DeepL Translator [2] are the current top 2 best machine translation tools using Transformer as the core architecture. They also support text-to-text, speech-to-text, and document-to-document cross-linguistic translation.
- More possible directions include low-resource language translation like Maori-to-other language and other language-to-Maori.

THE UNIVERSITY OF AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND

[1] https://ai.googleblog.com/2020/06/recent-advances-in-google-translate.html
[2] https://www.deepl.com/en/blog/how-does-deepl-work

Chatbot and Personal Mobile Assistant

- Personal mobile assistant and chatbots like XiaoIce, Siri, Google assistant, IBM watson are the applications that considering natural language processing and question answering which can answer user's question and chat with user.
- Moreover, medical chatbots like Mandy [1] can be used to help patients improve the efficiency of medical treatment, the patient can talk to the chatbot first, and then the chatbot collects the required user information according to a pre-defined template, and then summarize it into an electronic record. It can help doctors improve the medical treatment and simplify the process.
- Combining deep learning and knowledge graphs, we build a retrieval-based medical chatbot platform called HHH [2], and design an attention-based medical sentence similarity comparison algorithm, which can effectively improve the calculation of medical text similarity.

[1] L Ni et al. 2017. Mandy: Towards a smart primary care chatbot application. International symposium on knowledge and systems sciences.
[2] Bao, Q. 2020. HHH: an online medical chatbot system based on knowledge graph and hierarchical bi-directional attention. ACSW.

COVID-19 with NLP

- This study [1] focuses on predicting COVID-19 patient shielding — identifying and protecting patients who are clinically extremely vulnerable from coronavirus. This study focuses on techniques used for the multi-label classification of medical text. Using the information published by the United Kingdom NHS and the World Health Organisation, this paper present a novel approach to predicting COVID-19 patient shielding as a multi-label classification problem.

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

[1] Yogarajan, V. 2022. Predicting COVID-19 Patient Shielding: A Comprehensive Study. AJCAI.
[2] https://medium.com/@vithya.yogarajan/predicting-covid-19-patient-shielding-a-multi-label-classification-approach-6c517178e01d

References

- Bao, Q. et al., 2022. Multi-Step Deductive Reasoning Over Natural Language: An Empirical Study on Out-of-Distribution Generalisation. IJCLR-NeSy.
- Brown, T. et al. 2020. Language Models are Few-Shot Learners. NIPS 2020.
- Cho, K. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. EMNLP.
- Chung, J. et al. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, NIPS 2014.
- Cingillioglu, N. et al., 2018. DeepLogic: Towards End-to-End Differentiable Logical Reasoning, AAAI-MAKE19.
- Clark, P., et al. 2020. Transformers as Soft Reasoners over Language, IJCAI 2020.
- Devlin, J. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL 2018.
- Hartill, T., CONCEPTRULE, https://drive.google.com/file/d/1lxoAvtcvqVCYiO8e3tENnrTQ1NNVtpjs/view
- Hartill, T., CONCEPTRULE V2, https://drive.google.com/file/d/1lOCbW8bfZxj1RIzKDxn8xKg99XyYNj7z/view
- Hochreiter, S. et al. 1997. LONG SHORT-TERM MEMORY. Neural computation, 9(8), 1735-1780.

References

- Hudson, et al. 2018. Compositional attention networks for machine reasoning, ICLR.
- J J Hopfield et al. Neural networks and physical systems with emergent collective computational abilities. PNAS 1982.
- Kumar, et al. 2016. Ask me anything: Dynamic memory networks for natural language processing, ICML
- L Ni et al. 2017. Mandy: Towards a smart primary care chatbot application. International symposium on knowledge and systems sciences.
- Liu, Y. et al., 2019. Roberta: A robustly optimized bert pretraining approach. arxiv.
- Mikolov, T. et al. 2013. Efficient estimation of word representations in vector space.
- Young, N. et al. 2022. AbductionRules: Training Transformers to Explain Unexpected Inputs. The finding of ACL.
- Pennington, et al., 2014. Glove: Global vectors for word representation, EMNLP.
- Peter F. Brown et al. 1990. A Statistical Approach to Machine Translation. Comput. Linguistics, 16, 79-85.
- Programming in Prolog: Using the ISO standard, Clocksin, 2012

THE UNIVERSITY OF AUCKLAND
Te Whare Wananga o Tamaki Makaurau
NEW ZEALAND

References

- Qiu, X. et al. 2020. Pre-trained Models for Natural Language Processing: A Survey, Science China Technological Sciences, 63(10), 1872-1897.
- Radford, A. 2019. Language Models are Unsupervised Multitask Learners. OpenAI blog.
- Warren Weaver. 1955. Translation. Machine Translation of Languages, 14:15-23, 1955.
- Xiong, C., et al., 2016. Dynamic Memory Networks for Visual and Textual Question Answering, ICML.
- Vaswani, A. et al. 2017. Attention is all you need. NIPS.
- Yogarajan, V. 2022. Predicting COVID-19 Patient Shielding: A Comprehensive Study. AJCAI.

Thank you for your Attention!