AUTOMATIC PSYCHOLOGICAL TEXT ANALYSIS

JIMMY LAM

INTRODUCTION

Automated diagnoses of mental illness using schema therapy leads to faster and better recovery.

Use chat bots to predict schemas

Collecting labelled data is expensive and sensitive

Generate training data using AI

RESEARCH QUESTION

How well can a generative algorithm (e.g. RNN based encoder-decoder network) write stories that fit specific schemas?

- Most effective generative algorithms?
- Implementation and optimisations?
- Evaluation and comparison with Allaart's data

TU Delft

ALGORITHMS

1. OpenAI GPT
Transformer model that is pre-trained vast amounts of data and can beat the state of the art in NLP.

2. Recurrent neural networks
Neural network that exceeds in predicting results of sequential data because of its internal memory.

3. Generative adversarial networks
Minimax based model that pits two neural networks against each other to generate the best results.

Best candidate
OpenAI GPT is easy to use, is pre-trained and beats other models in 7 out of 8 times during zero-shot NLP tasks.

METHOD

Data set partitioning Removal of irrelevant entries	Pre-processing
Variant : 124k parameters Temperature : 0.7 / 1.0 Top_k : 0.0 Length : 250	Generation
Accomodate for overestimation and underestimation	Post-processing

RESULT

"i had a wonderful day today because my dads health was good, it lifted my spirits and i felt calm after a fewdays. i would say i was the happiest person i have ever been in a few days because of all the support i had received and i feel grateful to him for"

Coherence: 6/6 Correctness: 4/6

Schema correctness

	Conditional		Unconditional	
	is_happy	is_angry	is_happy	is_angry
Samples	63	59	63	59
C1 + C2	37	40	4	6
I1 + C2	9	7	3	4
C1 + I2	2	3	45	42
I1 + I2	24	9	10	7

Story independence

	Conditional		Unconditional	
BLEU	Is_happy	Is_angry	Is_happy	Is_angry
1-gram	0.16	0.09	0.2	0.12
2-gram	5.90e-155	4.06e-155	6.84e-20	1.24e-43
3-gram	4.80e-204	3.58e-204	2.22e-102	5.98e-93
4-gram	1.13e-231	8.84e-232	3.12e-112	9.27e-100

CONCLUSION

As sentences get longer the similarity of the generated samples decreases.

Generated stories can be assigned to a label with an accuracy of 58.7% or higher.

Post-processing and conditional prefixes are needed for an actual use case.

FUTURE WORK

Further analysis in OpenAI GPT-2 versions using a classifier.

Generation of stories with multiple schema's.