Natural language processing for cognitive therapy: Extracting schemas from thought records

Burger et al., *PLOS ONE*, 2021

Elijah Rockers, Cameron Buster

Overview

Concepts in cognitive therapy

Hypotheses

- Result comparison
- Difficulties

Reproducibility

Schemas

- Cognitive therapy attempts to change a patient's negative views on themselves, the world, or their future.
- This concept is known technically as "schemas"
- There are 9 core schemas as defined by Millings and Carnelly, 2015.
- *Attachment, Competence, Global self-evaluation, Health, Power and Control, Meta-cognition, Other people, Hopelessness, Other's views on self*

Thought records

Thought records are a way to collect free-text data in the form of *utterances*, or short sentences, that reflect underlying schemas.

Utterances were preprocessed by lower-casing, handling misspellings, contractions, and numbers, etc.

Normalized 100 dimensional GLoVE embeddings (circa 2014 total English Wikipedia articles) were used to represent the words in utterances, which were then represented as an average of word-vectors.

Thought record example

Table 1. Example of one complete thought record from the dataset collected in this study.

TR Question	Entry Type	Participant Response
Describe the situation very briefly in your own words.	open text entry field	while walking down the street I see someone I know, wave at them and they don't acknowledge my wave.
How well can you imagine yourself in this situation?	slider from 0 (not at all) to 100 (as good as if you were in the situation at the very moment)	85
Describe your emotion in this situation in one word.	open text entry field	disappointment
How intensely would you be experiencing this emotion?	slider from 0 (a trace) to 100 (the most intense possible)	45
Which of the following four emotions corresponds best with the emotion that you wrote down above?	multiple choice: sadness, fear, anger, happiness	sadness
Which (automatic) thought might have caused you to feel this way in the described situation?	open text entry field	They don't like me enough to wave back
And why would it be upsetting to you if "They don't like me enough to wave back" were true? What would it mean to you? What would it say about you?	*open text entry field*	*I may be unlikeable.*
And why would it be upsetting to you if "I may be unlikeable" were true? What would it mean to you? What would it say about you?	*open text entry field*	*I want friends. I will be lonely otherwise.*
And why would it be upsetting to you if "I want friends so I won't be lonely." were true? What would it mean to you? What would it say about you?	*open text entry field*	*If I am unlikeable then I won't have friends and will be alone all my life.*
What would you do in the situation, if anything?	open text entry field	I would try to make better impressions on people I meet.

Scoring utterances

Manual labeling was done for utterances, with each utterance being given a rating from 0 - 3.

I didn't do very well at dieting. => [0;3;0;2;0;0;0;0;0]

- Attachment, ***Competence***, Global self-evaluation, ***Health***, Power and Control, Meta-cognition, Other people, Hopelessness, Other’s views on self

Hypotheses

H1: NLP could be used to interpret thought records and form a determination of the underlying schema that may inform them.

H2: The DAT converges, meaning the deeper the thought record, the more likely the NLP method is to correctly predict the schema.

H4: There will be a correlation between predicted active schemas and the outcomes of several mental health questionnaires. (**Not reproduced**)

H1: SVM Regression Prediction

Attach	0.653487
Comp	0.619242
Global	0.431378
Health	0.335199
Control	0.236330
MetaCog	0.069315
Others	0.143586
Hopeless	0.506693
OthViews	0.479203

Schema-wise

Label = [0;3;0;2;0;0;0;0;0]

Prediction = [0;2;0;1;0;0;0;3;0]

Our Data

	KNN-C	KNN-R	SVM-C	SVM-R
Attach	0.56	0.58	0.60	0.65
Competence	0.66	0.64	0.67	0.62
Global Self Eval.	0.36	0.43	0.36	0.43
Health	0.77	0.54	0.65	0.34
Power and Ctl.	0.11	0.26	NaN	0.24
Meta-cog.	NaN	0.05	NaN	0.07
Other people	0.32	0.19	NaN	0.14
Hopelessness	0.50	0.43	0.42	0.51
Other Views	0.48	0.47	0.42	0.48

Their Data

Schema	Model Outcome					
	kNN-C	kNN-R	SVM	SVR	per-schema RNNs	multi-label RNN
Attachment	0.55 [0.51,0.60]	0.63 [0.59,0.65]	0.65 [0.61,0.68]	0.68 [0.65,0.70]	**0.73 [0.70,0.76]**	0.67 [0.66,0.72]
Competence	0.69 [0.64,0.73]	0.66 [0.63,0.69]	0.68 [0.65,0.72]	0.64 [0.61,0.67]	**0.76 [0.72,0.79]**	0.66 [0.64,0.69]
Global self-evaluation	0.40 [0.33,0.46]	0.41 [0.36,0.46]	0.36 [0.31,0.40]	0.49 [0.45,0.52]	**0.58 [0.54,0.63]**	0.49 [0.45,0.53]
Health	0.74 [0.65,0.81]	0.53 [0.44,0.60]	0.73 [0.65,0.81]	0.35 [0.31,0.40]	**0.75 [0.65,0.82]**	0.35 [0.31,0.39]
Power and Control	0.11 [0.02,0.18]	0.23 [0.17,0.27]	nan [0.00,1.00]	0.31 [0.26,0.35]	0.28 [0.20,0.35]	**0.31 [0.27,0.34]**
Meta-cognition	nan [0.00,1.00]	0.10 [0.01,0.20]	nan [0.00,1.00]	0.11 [0.06,0.16]	-0.01 [0.00,-0.01]	**0.11 [0.06,0.14]**
Other people	0.28 [0.00,1.00]	0.24 [0.17,0.31]	nan [0.00,1.00]	0.19 [0.14,0.24]	**0.22 [0.07,0.33]**	0.16 [0.10,0.20]
Hopelessness	0.48 [0.44,0.55]	0.51 [0.47,0.56]	0.49 [0.43,0.53]	0.54 [0.51,0.57]	**0.63 [0.56,0.68]**	0.53 [0.50,0.56]
Other's views on self	0.45 [0.41,0.51]	0.46 [0.42,0.50]	0.48 [0.43,0.53]	0.52 [0.48,0.55]	**0.58 [0.52,0.63]**	0.50 [0.47,0.54]

Difficulties

RNN models proved difficult to reimplement.

MLM-RNN: Gave results that differed wildly from the KNN and SVM methods, with very low goodness of fit scores.

PSM-RNN: Had bugs. 16 computational time made it difficult to solve this.

H2

Does the downward arrow converge?

Correlations for each depth: [0.59, 0.52, 0.55, 0.52, 0.67, 0.58, 0.72, 0.50, 0.49, 0.67, 0.65, 0.60]

#n for depth1 = 233, #n for depth 6 = 26, #n for depth 12 = 1

To determine if there was any correlation between depth and goodness of fit, we used a linear regression. As expected the r-value was low, and not significant.

Their H2 (excerpt):

The mean correlation between the predicted schema scores and the manually labeled schema scores was found to be 0.75 ($b = 0.75$, $t(220.76) = 46.97$, $p < 0.001$) when the nesting structure of utterances nested within thought records and thought records nested within participants is taken into account via random intercepts. Steps at a deeper level could not be scored better by the best model of H1 than steps at a more shallow level. The scoring accuracy, as measured by the Spearman correlation, did not improve with additional steps of the downward arrow technique ($\chi^2(1) = 1.21$, $p = 0.27$).

Difficulties

The original code was designed to use the PSM-RNN models to determine the results for H2.

The original code relied on implicit indexing of the data.

Code had to be written to re-associate utterances with the depth, as depth was never imported into the provided code. (*See h2.py*)

Reproducibility

- Exceptional availability of pre-processed data
- Bugs in provided code
- H2, H3, and H4 are not very well represented in the provided code, in particular H3 and H4.