Conceptual Reasoning for Generating Automated Psychotherapeutic Responses

Graham Mann(✉), Beena Kishore, and Pyara Dhillon

Murdoch University, 90 South Street, Murdoch, WA 6150, Australia
{g.mann,b.kishore,p.dhillon}@murdoch.edu.au

Abstract. The need for software applications that can assist with mental disorders has never been greater. Individuals suffering from mental illnesses often avoid consultation with a psychotherapist, because they do not realize the need, or because they cannot or will not face the social and economic consequences, which can be severe. Between ideal treatment by a human therapist and self-help websites lies the possibility of a helpful interaction with a language-using computer. A practical model of empathic response planning for sentence generation in a forthcoming automated psychotherapist is described here. The model combines emotional state tracking, contextual information from the patient's history and continuously updated therapeutic goals to form suitable conceptual graphs that may then be realized as suitable textual sentences.

Keywords: Natural language generation · Conceptual graphs · Model-based reasoning

1 Introduction

Many parts of the world now face a serious mental health care treatment gap, especially in low to middle income countries, and non-urban areas in high income countries [1]. The reasons are complex, but much of the shortage is caused by a lack of available skilled psychiatric professionals, and a failure of engagement by patients for economic or social stigma reasons [2]. A review of evidence shows that there are good reasons to think computerized therapy may be one effective approach to overcoming these difficulties [3]. While we do not imagine that these would be equivalent to consultation with skilled human psychiatrists, even existing mental health care apps can play a role and would often be better than nothing. In the case of "talking" therapies – those relying primarily on psychiatric interviews - software can today carry out natural conversations with a patient, simulating the role of the therapist. This paper deals with the formation and expression of appropriate responses to be used by an automated therapist during a consultation. It is a conceptual graph (CG) based language theory realized as a computer model of language generation.

Current trends in conversational systems tend to favour machine learning (ML) approaches, typically employing neural networks (NN), but we believe that these are not ideal in this application, for the following reasons. First, the knowledge and executable

M. Alam et al. (Eds.): ICCS 2020, LNAI 12277, pp. 186–194, 2020.
https://doi.org/10.1007/978-3-030-57855-8_14

skills of a machine learning system are typically opaque, lack auditability and so lack trust [4]. This is a serious drawback in medical applications. Knowledge and skills in conceptual graph (CG) based systems are as a rule much more human-readable and subject to logical reasoning that can readily be comprehended and verified. Second, NN-based or statistical ML approaches (with the possible exception of Bayesian learners) cannot easily incorporate high level, *a priori* knowledge into their processing [5]. This disadvantages learners in domains where such high-level knowledge is available or must be policy. But by virtue of their standardized knowledge representation, CG systems can freely mix prior knowledge incoming data relatively easily. Third, ML language systems are typically very data-hungry, and while large corpuses of language knowledge are now available, using these is computationally expensive. By contrast, model-based CG systems can, with some labour, be made to work with a relatively small amount of domain-specific language knowledge and with little or no learning.

In the rest of this paper, Sect. 2 proposes a linguistic model that integrates tracked emotional states, patient's utterances and background information on the patient with pragmatic cues from a control executive to generate a suitable response in conceptual form. Section 3 then describes our experimental implementation, consisting of heuristic Lisp functions to fetch instances of the above informative content from diverse sources, and calling on conceptual functions to bring these together to form CGs that can be realised as linear texts. The whole process is controlled by an expert system implementing psychotherapeutic rules. Finally, Sect. 4 concludes with some current challenges of this approach and its prospects for further development.

2 Sources Informing the Generation of Responses

2.1 Tracking of Patient's Expressed Emotions

It is difficult to imagine a successful psychotherapist who is not concerned with the emotional state of the patient. Even behaviourist therapies that emphasise overt actions in response to stimuli over mental state today include emotions as a recognised behavioural response, if not an important internal state determining them [e.g. 6]. The evidence is clear that the patient's emotional state is important for treatment needs to be closely monitored [7]. This state must be dealt with properly to maintain patients a comfortable place, while at the same time empathizing, noting the significance of the emotion and helping the patient to find meaning from it. Much emotional information can be obtained by monitoring a speaker's tone of voice, facial expression or other body language. Today's mobile devices, with their microphones and cameras could hope to read these forms of expression, but since our larger system is working with text alone it may not depend on non-verbal cues.

According to the survey conducted by Calvo and D'Mello [8] on models of affect, early approaches to detect emotional words in text include lexical analysis of the text to recognize words that are imminent of the affective states [9] or specific semantic analyses of the text based on an affect model [10]. Another approach is to construct affective models from large corpora of world knowledge and apply these models [e.g. 11]. The current work adapts Smith & Ellsworth's six-dimensional model [12] to make a system that can better grasp the subtleties of patient affect. Their chosen modal values on the principle component states for 15 distinguished emotional states are shown in Table 1.

One way that emotional tracking can be used is for the appropriate application of sympathy. We define a "safe region" in the 6D affective space. The therapist should be able to continue the therapy as long as the patient's tracked emotional state is within the safe region. A single point in the 6D affective space was chosen as the "most distressed" emotional state (we used {1.10 1.3 1.15 1.0 −1.15 2.0}). The simplest model of a safe region is outside a hypersphere of fixed radius centered on this point. The process is then reduced to finding the Euclidian distance between the current emotional state and the above-defined distressed center:

$$\Delta\Omega = \sqrt{(\mathcal{P}\mathrm{i} - \mathcal{P}\mathrm{j})^2 + (E\mathrm{i} - E\mathrm{j})^2 + (C\mathrm{i} - C\mathrm{j})^2 + (A\mathrm{i} - A\mathrm{j})^2 + (R\mathrm{i} - R\mathrm{j})^2 + (O\mathrm{i} - O\mathrm{j})^2}$$

Table 1. Mean locations of labelled emotional points in the range [−1.5, +1.5] as compiled in Smith & Ellsworth's study.

Emotion	P	R	C	A	E	O
Happiness	−1.46	0.09	−0.46	0.15	−0.33	−0.21
Sadness	0.87	−0.36	0	−0.21	−0.14	1.15
Anger	0.85	−0.94	−0.29	0.12	0.53	−0.96
Boredom	0.34	−0.19	−0.35	−1.27	−1.19	0.12
Challenge	−0.37	0.44	−0.01	0.52	1.19	−0.2
Hope	−0.5	0.15	0.46	0.31	−0.18	0.35
Fear	0.44	−0.17	0.73	0.03	0.63	0.59
Interest	−1.05	−0.13	−0.07	0.7	−0.07	0.41
Contempt	0.89	−0.5	−0.12	0.08	−0.07	−0.63
Disgust	0.38	−0.5	−0.39	−0.96	0.06	−0.19
Frustration	0.88	−0.37	−0.08	0.6	0.48	0.22
Surprise	−1.35	−0.97	0.73	0.4	−0.66	0.15
Pride	−1.25	0.81	−0.32	0.02	−0.31	−0.46
Shame	0.73	1.31	0.21	−0.11	0.07	−0.07
Guilt	0.6	1.31	−0.15	−0.36	0	−0.29

If the calculated distance is greater than an arbitrarily-defined tolerance threshold (radius), the patient's current emotional state is considered safe. The calculated ΔΩ of an emotional state {1.15 0.09 1.3 0.15 −0.33 −0.21} from the above-defined distress point would be 1.70. For an arbitrary tolerance radius of 2.5 units from the distress point, the patient's tracked emotive state would not be in the safe region. Further work on a better model of the actual "shape of distress" might improve the heuristic's ability to pick a highly appropriate response for any given emotional state.

2.2 Conceptual Analysis of Patient's Utterances

Study of a reference corpus of 118 talking therapy interviews [13], reveals that these patient utterances can be long and rambling, often incoherent and quite difficult for a person, much less a machine, to comprehend. While we have a conceptual parser, SAVVY, capable of converting real, non-grammatical paragraphs into meaning-preserving CGs [14], it was not developed for use in this domain. Though possible in principle, for the present work we do not intend to improve it to the point of creating meaningful conceptual representations for most of the utterances observed in our corpus. Conceptual parsers depend on an ontology in the form of a hierarchy of concepts, a set of relations and a set of actors. Manually creating representations of all the terms used in those interviews for SAVVY would be a very difficult and time-consuming task. (This most serious of drawbacks for conceptual knowledge-based systems is now being addressed in automated ontology-building machines [e.g. 15, 16]). Our focus in this study is the *generation* of language.

Yet this kind of psychotherapy is essentially conversational, so we must allow the conceptual representations of patient utterances to be an input even to test response formation. Therefore, SAVVY will be adapted to accept selected patient utterances of interest. In some cases, to keep the project manageable, we hand-write plausible input CGs to avoid diverting too much time and energy away from our generation pipeline.

2.3 Using Context to Inform the Planning Process

In regular clinical practice, the first step for a new patient is an admitting (or triage) interview, that can capture important biographical details, a presenting complaint, background histories, and perhaps an initial diagnosis. Because we wish our model of language generation to account for existing, contextual information, we will not actively model this initial interview, but rather only subsequent interviews that have access to this previously gathered background. A set of background topics that should be sought during an admitting interview is described by Morrison [17]. Our current model draws 12 topics from this source and adds three extra topics specific to our clinical model.

2.4 Executive Control

An executive based on a theory about how therapy should be done is needed for overall control. At each conversational turn, the executive should recommend the best "pragmatic move" for the response generation process. This allows selection and instantiation of appropriate high-level conceptual templates that form the therapist's utterances to support, guide, query, inform or sympathize with the patient as appropriate during the treatment process. Our executive is based on the brief therapy of Hoyt [18] and the solution-based therapy of Shoham et al. [19]. As recommended by Hoyt, the focus is on negotiating treatment practices, not diagnostic classification. However, in this experiment a working diagnosis might become available as a result of the therapy or be input as background knowledge.

For a natural interviewing style, the executive must allow its goal-seeking behaviour to be interrupted by certain imperatives imposed by conversational conventions and good clinical practice. If the patient asks a question, this deserves some kind of answer. If the patient wishes to express some attitude or feeling about some point, that should usually be entertained immediately. If the patient's estimated emotional state falls into distress, it is important that the treatment model is suspended until the patient can be comforted and settled. Similarly, if rapport with the patient is lost (the quality of the patient's responses deteriorates), special steps must be taken to recover this before anything else can be done. We call these *forced* responses, to distinguish them from less obligatory pragmatic moves, which in our model are driven by key goals in the therapy.

In most cases, a conceptual structure representing a suitable therapist's response can be formed by unifying pragmatically-selected schemata with content-bearing information from the other sources. This process is to be handled by heuristic rules that must be sufficiently general to keep the number needed as low as possible. In a few cases, a single standardized expressive form can be accessed without the need for unification.

3 Implementation Details

3.1 Collection of Emotional State, Patient Utterances and Background Knowledge

To track emotions, we are experimenting with computationally cheap heuristics that can distinguish the patient's current emotional states directly from the text (Fig. 1), though this has the disadvantage that it does not model cognitive aspects of emotion. For example, a patient would be talking about his or her current emotional reaction if there is a cluster of words in an utterance that includes, within a window of *n* words, one or more members of the word bag {I, I'm, me, myself}, optionally one of {feel, think, consider, say} AND at least one emotion cue such as {hate, love, enjoy, relax…etc.}. Additional rules have to be implemented to account for negation and to check for past tense such as {was, did, suffixes such as 'ed'}. The 15 emotional states were annotated with synonyms retrieved from WordNet-Affect [20]. A short list of common generic non-emotional words such as {the, then, who, when… etc.} is also provided (excluded_list).

```
Function DetectEmotion (text)
 {For each word in a text
   {If the word is in one of the 15 annotated lists
      {If the emotion belongs to the patient and not past tense
         {emotion_list ← push the 6D value of identified state
         }}}
   {If #(emotion_list) > 1
       emotional_state ← mean(emotion_list)}
    Else {emotional_state ← emotion_list}
return emotional_state  }
```

Fig. 1. Pseudocode for the heuristic function detecting the emotional state in the utterance

To bring patient's conversational utterances into the picture, a text-to-CG parser is required. But even if it was feasible to construct complete representations for every utterance made by a patient, this would not be desirable, because from analysis of the corpus, surprisingly few such representations would actually have useful implications for treatment, at least within our simplified model. Therefore whatever method we use to parse patient input into CGs can afford to be selective about the outputs it forms, using top-down influences to prefer those interpretations that are likely to lead to useful content. Our conceptual parser, SAVVY, can do this because it assembles composite CGs out of prepared conceptual components that are already pre-selected for the domain of use to which they will be put. This means that the composite graphs are strongly weighted toward semantic structures of implicit value, leaving utterances that result in disjoint or useless component parts not able to aggregate at all. For example, if the patient says

I'm scared that one day, he'll just stab me

SAVVY is able to assemble the conceptual graph

```
[FEAR] –
        (expr) -> [PERSON:Patient]
         (attr) -> [NEGATIVE]
        (caus) ->[ (futr) -> [SITUATION: [PERSON] < - (agnt) <- [PTRANS] -
                                                        (ptnt) -> [PERSON: Patient] - - - - - - - - - - - - - - - - - -¡
                                                         (rslt) -> [PHARM] -> (expr ) -> [PERSON:Patient] - -!
                                                         (inst) -> [BLADE]
                                                        (ptim) -> [DAY:@1]  ]
```

only because the lexicon had definitions of useful subgraphs for "scared", "stab" etc. which, because they would likely represent harm to someone, are considered important inclusions. Not every act or even every emotion is so provided for.

A simple database currently provides background knowledge for our experiments. Each entry in the knowledgebase is a history list of zero or more CGs, indexed by both a patient identifier and one of the 15 background topics (Sect. 2.3) such as suicide_attempts, willingness_to_change and chief_complaint. Entries may be added, deleted or modified during processing, so the database can be used as a working memory to update and maintain therapeutic reasoning over sessions. Initially these entries are provided manually to represent information from the pre-existing admitting interview, but these can be updated, edited or deleted by the automated therapist. Automatic entries are vetted by domain-specific heuristic filters focussed on the topic of interest, so that only relevant CGs can be pushed onto the appropriate lists. Our experience suggests that it is not difficult to write these provided high-level conceptual functions, based on the canonical operators, are available to find and test specific sections of the graphs.

3.2 Expert System for Executive Control

Psychiatric expertise is represented by a clinical Expert System Therapist (EST), based on TMYCIN [21]. This shell is populated with rules from the above-mentioned treatment theories. Consultation of the system is performed at each conversational turn, informed by the current state of variables from the inputs. Backward-chaining inference maintains internal state variables and recommends the best "pragmatic move" and" therapist's target" for the response generation process. These parameters allow the selection and instantiation of appropriate high-level template graphs that form the therapist's utterances to support, guide, query, inform or sympathize at that moment. In some simple cases, canned responses are issued to bypass the language pipeline and reduce processing demands.

3.3 Response Generation

Responses CGs are generated based on the three input sources and two variables from the therapeutic process (Fig. 2). The heuristic first checks the patient's emotional state. If this is outside the distressed region (Sect. 2), an expressive form CG is created by first maximally joining generalised CG templates for the pragmatic move (e.g. query) and the content recommended for the current therapeutic goal (e.g. establish_complaint). This is then instantiated with background information. If the patient's emotional state is inside the distressed region, the EST will recommend a pragmatic move of sympathy and the heuristic will force the use of a sympathy template for its expressive form. The constructed CG is subsequently passed to a realization heuristic, SentenceRealization() which in turn expresses the CG as a grammatically correct sentence using YAG (Yet Another Generator) [22].

```
Function GenerateSentence (emotional_state, patient_CG,  background_info, pragmatic_move, therapeutic_goal) {
  If the emotional_state is not in distressed region {
     pragmatic_move_CG ← retrieve the template for this pragmatic move
     therapeutic_goal_CG ← retrieve the CG for this therapeutic goal, elaborated from patient_CG
     expressive_form_CG ← maximal_join (pragmatic_move_CG, therapeutic_goal_CG)
     constructed_CG ← instantiate_background (expressive_form_CG, background_info)
     SentenceRealization (constructed_CG, pragmatic_move)}}
  else {
 SentenceRealization (sympathy_CG, pragmatic_move)}
```

Fig. 2. Pseudocode for the heuristic function creating the expressive form.

YAG is a template-based syntactic realization system, which comes with a set of core grammar templates that can be used to generate noun phrases, verbs, prepositional phrases, and other clauses. We are developing a further set of custom templates using those core grammar templates with optional syntactic constraints. The heuristic function *SentenceRealization*()(Fig. 3) loops through the constructed CG and creates a list of attribute-value pairs, based on grammatical/semantic roles.

```
Function SentenceRealization (constructed_CG, pragmatic_move)
{For each concept in a CG
     {knowledge_list ← push attribute-value pair onto knowledge_list}
  Retrieve template based on the pragmatic move and the constructed_CG
  For each slot in the template
     {Override the slot default value with the value from knowledge_list}
  Realize the template using the command surface-1 of YAG.}
```

Fig. 3. Pseudocode for realizing the expressive form as text.

4 Conclusion

This generation component is still in development, so no systematic evaluation has yet been conducted. Some components have been coded and unit tested. Getting the heuristics of the system to interact smoothly with each other is a challenge; that is to be expected in this modelling approach. We are concerned about the number of templates that may be required, particularly at the surface expression level. If they become too difficult or too many to create, the method might become infeasible. The heuristic tests are not difficult to write, but are, of course, imperfect compared to algorithms. Also, we have not fully tested the emotion tracking on real patient texts so far.

Our planned evaluation has two parts. First, a systematic "glass-box" analysis will discover the strengths and limitations of the generation component, particularly with respect to the amount of prior knowledge that needs to be provided and the generality of the techniques. Second, the "suitability", "naturalness" and "empathy" of the response generation for human use will be tested, using a series of ersatz patient interview scenarios to avoid the ethical complications of testing on real patients. The scenarios will provide human judges (expert psychotherapists or, more likely, students in training to be psychotherapists) with information about an ongoing therapeutic intervention. Example patient utterances and the actual responses generated by the system will also be provided as transcripts. The judges will then rate these transcripts on those variables using their own knowledge of therapy.

Finally, we reiterate that if hand-built conceptual representations can be practically built up using existing methods, the effort will be worthwhile if the systems are then more transparent and auditable than NN or statistical ML system and thus, more trustworthy.

References

1. Jack, H.E., Myers, B., Regenauer, K.S., Magidson, J.F.: Mutual capacity building to reduce the behavioral health treatment gap globally. Adm. Policy Mental Health Mental Health Serv. Res. **47**(4), 497–500 (2019). https://doi.org/10.1007/s10488-019-00999-y
2. Meltzer, H.E., et al.: The reluctance to seek treatment for neurotic disorders. Int. Rev. Psychiatry **15**(2), 123–128 (2003)
3. Fairburn, C.G., Patel, V.H.: The impact of digital technology on psychological treatments and their dissemination. Behav. Res. Therapy **88**, 19–25 (2017)
4. Marcus, G.: Deep learning: a critical appraisal. arXiv preprint arXiv:1801.00631 (2018)
5. Pearl, J.: Theoretical impediments to machine learning with seven sparks from the causal revolution. arXiv preprint arXiv:1801.04016 (2018)
6. Ellis, A.: Rational-emotive therapy. Big Sur Recordings, CA, USA, pp. 32–44 (1973)
7. Greenberg, L.S., Paivio, S.C.: Working with Emotions in Psychotherapy, vol. 13. Guilford Press, New York (2003)
8. Calvo, R.A., D'Mello, S.: Affect detection: an interdisciplinary review of models, methods, and their applications. IEEE Trans. Affect. Comput. **1**, 18–37 (2010)
9. Hancock, J.T., Landrigan, C., Silver, C.: Expressing emotion in text-based communication. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 929–932. Association for Computing Machinery (2007)
10. Gill, A.J., French, R.M., Gergle, D., Oberlander, J.: Identifying emotional characteristics from short blog texts. In: 30th Annual Conference of the Cognitive Science Society, Washington, DC, pp. 2237–2242. Cognitive Science Society (2008)

11. Breck, E., Choi, Y., Cardie, C.: Identifying expressions of opinion in context. In: IJCAI, vol. 7, pp. 2683–2688, January 2007
12. Smith, C.A., Ellsworth, P.C.: Attitudes and social cognition. J. Pers. Soc. Psychol. **48**(4), 813–838 (1985)
13. McNally, A., et al.: Counseling and Psychotherapy Transcripts, Volume II. Alexander Street Press, Alexandria (2014)
14. Mann, G.A.: Control of a navigating rational agent by natural language. Unpublished Ph.D. thesis, University of New South Wales, Sydney, Australia (1996). https://manualzz.com/doc/42762943/control-of-a-navigating-rational-agent-by-natural-language
15. Paola, P.V., et al.: Evaluation of OntoLearn, a methodology for automatic learning of domain ontologies. In: Ontology Learning from Text: Methods, Evaluation and Applications, vol. 123, p. 92 (2005)
16. Leuzzi, F., Ferilli, S., Rotella, F.: ConNeKTion: a tool for handling conceptual graphs automatically extracted from text. In: Catarci, T., Ferro, N., Poggi, A. (eds.) IRCDL 2013. CCIS, vol. 385, pp. 93–104. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-54347-0_11
17. Morrison, J.: The First Interview: A Guide for Clinicians. Guilford Press, New York (1993)
18. Hoyt, M.F.: The temporal structure of therapy. In: O'Donohue, W.E., et al. (ed.) Clinical Strategies for Becoming a Master Psychotherapist, pp. 113–127. Elsevier (2006)
19. Shoham, V., Rohrbaugh, M., Patterson, J.: Problem-and solution-focused couple therapies: the MRI and Milwaukee models. In: Jacobson, N.S., Gurman, A.S. (eds.) Clinical Handbook of Couple Therapy, pp. 142–163. Guilford Press, New York (1995)
20. Strapparava, C., Valitutti, A.: WordNet-affect: an affective extension of WordNet. In: 4th International Conference on Language Resources and Evaluation, pp. 1083–1086 (2004)
21. Novak, G.: TMYCIN expert system tool. Technical Report AI87–52, Computer Science Department, University of Texas at Austin (1987). http://www.cs.utexas.edu/ftp/AI-Lab/tech-reports/UT-AI-TR-87-52.pdf. Accessed 5 Feb 2018
22. Channarukul, S., McRoy, S.W., Ali, S.S.: Enriching partially-specified representations for text realization using an attribute grammar. In: Proceedings of the 1st International Conference on NLG, Mitzpe Ramon, Israel, vol. 14, pp. 163–170. Association for Computational Linguistics (2000)