

Annotating Reflections for Health Behavior Change Therapy

Nishitha Guntakandla *and* Rodney Nielsen

Human Intelligence and Language Technologies

University of North Texas

Denton, Texas 76201

nishithaguntakandla@my.unt.edu *and* Rodney.Nielsen@unt.edu

Abstract

We present our work on annotating *reflections*, an essential counselor behavioral code in motivational interviewing for psychotherapy on conversations that are a combination of casual and therapeutic dialogue. We annotated all the therapists' utterances from ten transcripts spanning more than five hours of conversation with Complex Reflection, Simple Reflection, or No Reflection. We also provide insights into corpus quality and code distributions. The corpus that we constructed and annotated in this effort is a vital resource for automated health behavior change therapy via a dialogue system. As the on-going work, additional conversations are being annotated at least by one annotator.

Keywords: Reflections, Motivational Interview, Behavioral codes, Companion robot conversations, Virtual health agents, Social-therapeutic agents

1. Introduction

Recent research has focused on developing companionable robots (companion bots) (Sarma et al., 2014) and relational agents (Vardoulakis et al., 2012) who build trust, rapport, and therapeutic alliance over time by engaging in both casual, and therapeutic dialogue. The goal of such social-therapeutic agents¹ is to play a key role in health care at home. They do so by incorporating therapeutic dialogue into friendly conversations with elderly individuals suffering from depression (Nielsen et al., 2010; Ring et al., 2016; Bickmore et al., 2005). Older adults those who lack social support can have negative effects on the health and well-being (Vardoulakis et al., 2012). Furthermore, it is not possible to have a therapist available to every individual at each instance of their life to provide assistance. Situations like these enhance the need for social-therapeutic agents.

As psychology studies (Keeley et al., 2014; Arkowitz et al., 2015) show, depression can be treated by Motivational Interviewing (MI) (Miller and Rollnick, 2012), a goal-oriented and client-centered style of psychotherapy. In MI, the therapist is directive in aiding the client to elicit the change talk, resolve ambivalence, and evoke motivation to make positive behavioral changes. The proficiency of the counselor in MI delivery is measured using behavioral coding systems like the MI Skills Code (MISC)(Miller et al., 2003) or the Motivational Interviewing Treatment Integrity (MITI)(Moyers et al., 2014). These systems are designed to manually annotate the counselor and the client utterances in a session with behavioral codes, which is an extremely labor intensive and costly task.

Existing work (Can et al., 2012; Pérez-Rosas et al., 2016; Tanana et al., 2015) annotated behavioral codes exclusively on MI sessions, which predominantly focus

on therapeutic dialogue. MI sessions are conducted to help clients change behaviors that they identify or agree to being very important to improve their health. Unlike MI sessions, social-therapeutic agent conversations are a combination of casual (non-therapeutic) and therapeutic dialogue. They commence the conversation with rapport building and identify a behavior that should be discussed. Hence, there is a need for building an appropriate corpus that can be used to develop such agents. We address this need by constructing and annotating companionbot² conversations with behavioral codes. In this work, as the first step, we only focus on annotating *reflections* among all the counselor behavioral codes .

Reflective listening is a core MI micro-skill that therapists possess to illustrate their interest in a client's declaration, which intensifies the client's trust in the therapist. Reflections are statements that define therapeutic presumptions about what the client is trying to express about their inner experience (Miller et al., 2002). Specifically, counselor reflections are the thoughts, feelings, and meaning of what a client said. Reflections tend to be collaborative and non-judgmental and aim to guide the client towards resolving ambivalence. The outcome of using reflective statements is the client feels listened to, understood, less resistant, and more motivated to change. Counselor reflections can be categorized into Simple or Complex. Simple Reflections (SR) typically add little or no meaning to what the client has said. They do not go far beyond the client's original statement. Complex Reflections (CR) add substantial meaning by using inference and synthesis to state something the client has implied. Following is an instance presenting these two forms of reflection:

Client: *My mother is driving me crazy. She says she wants to remain independent, but she calls me four times a day with trivial questions. Then she gets mad when I give*

¹social-therapeutic agents: This term is used to indicate any computer-based agent or bot that provide both social support and therapy.

²In our work we used companionbot, a specific type of social-therapeutic agent to collect the data

Therapist: So, they -- they pick you up and take you to their home, I guess?

Participant: Oh, yeah, mhm, yeah, I have an apartment at their place.

Therapist: Okay.

Participant: But {Breath} -- it's downstairs.
And, uh, it's -- I'm too much trouble for the, you know -- not on their end,
but I -- I feel I'd be just too much trouble staying there for them to take care of all the time.

Therapist: Okay, I see.
So where do you live right now?

Participant: Uh, in {Breath} -- in an assisted, uh, -- well an independent care apartment house.

Figure 1: Sample transcript from companionbot conversations.

her advice.

Interviewer: Things are very stressful with your mother. (SR)

Interviewer: You're having a hard time figuring out what your mother really wants. (CR)

In this work, we:

- Describe how we collected behavioral code annotations for more than 1536 counselor utterances taken from wizard-of-Oz-style conversations between human and companionbots.
- Conduct a corpus analysis, and present both inter-annotator agreement and code distribution among annotated utterances.



Figure 2: Prototype of our companionbot that is placed in the participants' room for them to interact with.

2. Corpus Construction

In this section, we describe the collection of companionbot conversations over which we annotate reflections.

2.1. Data Collection

We conducted Wizard of Oz experiments (Dahlbäck et al., 1993) to collect data for our corpus. Three psychologists (wizards) were employed to engage in one-to-one conversations with participants remotely. A total of 16 elderly individuals from the Denver, Colorado and Dallas-Fort Worth, Texas metro regions participated in these experiments. A fixed physical companionbot (shown in Figure 2) was placed in the participant's room for them to interact with at the scheduled times. Arrangements were made for a wizard to interact with a subject for 30 minutes in each session. Additionally, the psychologists were given guidelines on how to play the role of the companionbot and conduct appropriate dialogue. As a whole, 324 sessions were recorded. Audio files from all and video files from most of the sessions were gathered. The final corpus comprises a total of 162 hours of audio.

2.2. Transcription

The collected audio files were transcribed manually by linguists following the guidelines from (Strassel, 2003). The identifiers, *Therapist* and *Participant*, are assigned to the psychologist and the participant turns, respectively. On average, each transcript contains 115 turns of therapist and participant dialogue. Within each turn, a speaker produces multiple sentences/utterances. Speech is segmented at the end of a sentence. This sentence segmentation follows the directives:

1. Sentence is an independent clause.
2. If a complex sentence contains a coordinating conjunction (*and*, *but*, *for*, *nor*, *or*, *so*, and *yet*) and the following clause is not dependent on the first clause, then these two clauses are segmented into two separate sentences.

Also, additional rules are followed to deal with noise, pauses, and false starts.

Utterances (Behavioral Code)
<p>Participant: So I know that in the past, what I did was I either got down on the floor and did, let's say, twenty sit ups, {Breath} {Pause} or I walked around two blocks. And by the time you do anything physical like that, {Pause} you – you've forgotten it. You d – you no longer wanna go grab that food. {Pause} [At least I], Therapist: [Mhm] (NR) {Pause} (NR) Participant: At least I don't. It's over! The moment has passed. Therapist: Yeah. (NR) {Pause} (NR) Yeah. (NR) Exercise is a l- +lot – it's a lot better (CR) Participant: Yeah. Therapist: than – than eating. (CR)</p>
<p>Therapist: And how is your health right now? (NR) Participant: It's not the best. (()) but they say I'm getting along. (()) so that's what counts. Therapist: Yeah, so how old are you? (NR) Participant: Uh, eighty. Just turned eighty. Therapist: Okay, {Pause} so, your health is not the best (SR)</p>

Table 1: Example annotations illustrating Simple, Complex and No reflections.

3. Annotation

Our annotation process is presented in this section. To compute inter-annotator reliability, as the first stage our annotators annotated ten of the acquired transcripts (Section 2.2.). These transcripts cover more than five hours of conversation, and contain a total of 1536 counselor utterances. They were acquired from multiple participants and form a representative portion of the collected data set (Section 2.1.). Each transcript was annotated by at least two annotators, and a total of three annotators participated in this task. The annotators label each counselor utterance³ as: Complex Reflection (CR), Simple Reflection (SR), or No Reflection (NR). The remaining transcripts are being annotated by at least one of the three annotators (those annotations are still in progress). Table 1 depicts a sample annotation from the corpus. We describe the guidelines and tool employed for annotation below.

3.1. Guidelines

MISC and MITI are coding systems have been used extensively in prior work to manually annotate counselor and client utterances with behavioral codes. MISC is an exhaustive system that defines rules for assigning codes to both counselor and client utterances. MITI is built on MISC which solely targets counselor behavior. Since reflection is essentially counselor behavior, we use the MITI 4.2.1 standard (Moyers et al., 2014) as part of our guidelines while annotating reflections. Additionally, we define the following rules that supersede MITI to annotate our transcripts:

1. Utterances that are structured as reflections but, fall under a different behavioral code according to the MITI guidelines are also annotated as reflections. For instance, declarative questions and reflections are only distinctive with the tone in which the speech is delivered.
2. When a counselor provides a series of reflections in a single turn, all the utterances in the series are annotated as individual reflection.
3. If a client speech or noise interrupts a counselor's reflection, the utterance splits. In this case, both the utterances are coded with reflection.

3.2. Annotation Tool

We use GATE (General Architecture for Text Engineering) (Cunningham et al., 2013) to ease the annotation process and promote consistency. Annotation codes are enumerated in an *annotation schema*, an xml file. The GATE user interface is used by the annotators to map counselor utterances in the transcripts to the codes defined in the annotation schema. Finally, an output file enriched with annotations is generated.

4. Corpus Analysis and Quality

We analyze the code distribution, assess the quality of the annotated corpus, and present insights in this section. Table 3 presents the raw counts and percentage of code distribution annotated in the corpus. The total amount of reflection observed in the annotated transcripts is 8% (125) of total counselor utterances. Unlike companionbot

³In this work, we use the terms *sentence* and *utterance* interchangeably.

Work	Corpus	# Sessions	Duration	# Utterances	Agreement
(Can et al., 2012)	Drug abuse problem (HMCBI), alcohol use disorders (ESP21, ESPSB)	12	-	-	Overall: 0.661 - 0.764
(Pérez-Rosas et al., 2016)	Clinical Trails, Standardized patients, Brief MI encounters, Coaching phone calls	10	4.5 hrs	1160	CR : 0.49 SR : 0.34
Our Work	Conversational bot conversations	10	5 hrs	1536	Overall: 0.64 CR: 0.64 SR: 0.49

Table 2: Comparison of existing corpora annotated with behavioral codes with our work

conversations, MI corpora are entirely therapeutic and usually portray a higher percent of reflection.

Code	Count	Percentage
Complex Reflection (CR)	98	6.4%
Simple Reflection (SR)	27	1.7%
No Reflection (NR)	1411	91.9%

Table 3: Distribution of each code in the annotated corpus.

Code combination	Kappa
CR vs SR vs NR	0.64
(CR or SR) vs NR	0.69
CR vs (SR & NR)	0.64
SR vs (CR & NR)	0.49

Table 4: Inter-annotator reliability for different combination of codes expressed by Cohen's Kappa values.

We use Cohen's Kappa (Viera et al., 2005) to assess inter-annotator reliability in our corpus. Table 4 presents the Kappa values between several combinations of codes. The Kappa computed over whole annotated counselor utterances with three codes is 0.64 which is considered to be substantial agreement for categorical values (Landis and Koch, 1977). For a detailed analysis, we also report annotators' agreement on each code separately. Scores indicate that annotators identified any reflection (i.e., SR or CR), and only CR in transcripts, fairly well. We observed that three of the annotated transcripts contain zero SR and its overall frequency is low, which justifies the lower Kappa score obtained for SR.

To better perceive the agreements, we also present the normalized contingency matrix between the annotators for all codes in a heat map (Figure 3). High diagonal values in the heat map indicate that our annotations are reliable. The heat map also indicates minor ambiguity between CR and NR. This is because of the complexity involved in distinguishing the therapeutic dialogue from the non-therapeutic dialogue in some cases. For instance, consider the following sample turn from one of the annotated transcripts:

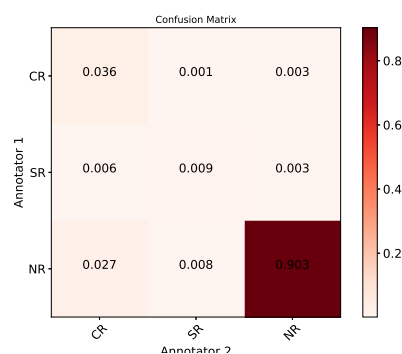


Figure 3: Heat map for the normalized contingency matrix between the labels of the two annotators.

Therapist: [Yes, you – you] said that he had children and the children live with him or the mom?

Participant: Uh, they stay split fifty-fifty.

Therapist: Okay, so they also get exposed to some of this trauma.

The sentence in the example of Therapist's last turn is a complex reflection in casual conversation.

5. Previous Work

Companionbots that can be used in health care at home need to engage in conversations that are both casual and therapeutic. Research on MI interventions that are solely therapeutic (Can et al., 2016; Pérez-Rosas et al., 2017; Tanana et al., 2015) is being actively pursued. Efforts focused on collecting MI session transcripts and annotating them with behavioral codes have become popular. In this section, we detail on how we expand on the existing resources by incorporating therapeutic dialogue within a non-therapeutic setting.

Mainly, we compare and contrast our novel annotated corpus of companionbot conversations with related corpora (Table 3.1.). (Can et al., 2012) is one of the prominent works that annotated both counselor and client utterances following MISC guidelines. This work reports inter-annotator agreement on 12 MI sessions with an overall agreement of 0.764 and 0.661 when considered at utterance-level and word-level, respectively. Authors mention the actual code agreement lies in between the

above specified scores. Moreover, (Pérez-Rosas et al., 2016) is another significant work that focused explicitly on annotating counselor behavioral codes following MITI guidelines. Inter-annotator agreement of 0.49 and 0.34 is reported for CR and SR codes, respectively. The characteristics of our annotated data set correspond with other existing corpora. Although it is comparatively difficult to detect the therapeutic reflections among casual and therapeutic utterances, it is evident from this comparison that our corpus is also a high quality indispensable resource.

The sensitive nature of Motivational Interviewing sessions do not permit the public release of the existing corpora in this area. However, we are constructing a corpus that can be distributed for on-going research.

6. Conclusion

In this work, we detail the collection of companionbot conversations and the process of annotating reflections on counselor utterances. The inter-annotator agreement indicates that the annotations have substantial quality. We exemplify any ambiguity present in the annotations and specify the reasons behind it. Our corpus addresses the need for a suitable corpus that would enable the development of companion bots and other dialogue systems focused on identifying and processing complex reflections in conversations with therapists. As the work in progress, at least one of the annotators are annotating the additional conversations. This corpus will be available for research upon request⁴.

7. Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant 1262860. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

8. References

- Arkowitz, H., Miller, W. R., and Rollnick, S. (2015). *Motivational interviewing in the treatment of psychological problems*. Guilford Publications.
- Bickmore, T. W., Caruso, L., Clough-Gorr, K., and Heeren, T. (2005). 'it's just like you talk to a friend' relational agents for older adults. *Interacting with Computers*, 17(6):711–735.
- Can, D., Georgiou, P. G., Atkins, D. C., and Narayanan, S. S. (2012). A case study: Detecting counselor reflections in psychotherapy for addictions using linguistic features. In *INTERSPEECH*, pages 2254–2257. ISCA.
- Can, D., Marín, R. A., Georgiou, P. G., Imel, Z. E., Atkins, D. C., and Narayanan, S. S. (2016). âit sounds like...â: A natural language processing approach to detecting counselor reflections in motivational interviewing. *Journal of counseling psychology*, 63(3):343.
- Cunningham, H., Tablan, V., Roberts, A., and Bontcheva, K. (2013). Getting more out of biomedical documents with gate's full lifecycle open source text analytics. *PLoS computational biology*, 9(2):e1002854.
- Dahlbäck, N., Jönsson, A., and Ahrenberg, L. (1993). Wizard of oz studiesâwhy and how. *Knowledge-based systems*, 6(4):258–266.
- Keeley, R. D., Burke, B. L., Brody, D., Dimidjian, S., Engel, M., Emsermann, C., Thomas, M., Morales, E., Koester, S., Kaplan, J., et al. (2014). Training to use motivational interviewing techniques for depression: a cluster randomized trial. *The Journal of the American Board of Family Medicine*, 27(5):621–636.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.
- Miller, W. R. and Rollnick, S. (2012). *Motivational interviewing: Helping people change*. Guilford press.
- Miller, W. R., Moyers, T. B., Ernst, D., and Amrhein, P. (2003). Manual for the motivational interviewing skill code (misc). *Unpublished manuscript. Albuquerque: Center on Alcoholism, Substance Abuse and Addictions, University of New Mexico*.
- Moyers, T., Manuel, J., and Ernst, D. (2014). Motivational interviewing treatment integrity coding manual 4.1. *Unpublished manual*.
- Nielsen, R. D., Voyles, R. M., Bolanos, D., Mahoor, M. H., Pace, W. D., Siek, K. A., and Ward, W. H. (2010). A platform for human-robot dialog systems research. In *AAAI Fall Symposium: Dialog with Robots*.
- Pérez-Rosas, V., Mihalcea, R., Resnicow, K., Singh, S., and An, L. (2016). Building a motivational interviewing dataset. In *CLPsych@HLT-NAACL*, pages 42–51. The Association for Computational Linguistics.
- Pérez-Rosas, V., Mihalcea, R., Resnicow, K., Singh, S., Ann, L., Goggin, K. J., and Catley, D. (2017). Predicting counselor behaviors in motivational interviewing encounters. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, volume 1, pages 1128–1137.
- Ring, L., Bickmore, T., and Pedrelli, P. (2016). Real-time tailoring of depression counseling by conversational agent. *Iproceedings*, 2(1):e27.
- Sarma, B., Das, A., and Nielsen, R. (2014). A framework for health behavior change using companionable robots. In *INLG*, pages 103–107. The Association for Computer Linguistics.
- Strassel, S. (2003). Six corpus of classical sociolinguistics interviews.
- Tanana, M., Hallgren, K., Imel, Z. E., Atkins, D. C., Smyth, P., and Srikumar, V. (2015). Recursive neural networks for coding therapist and patient behavior in motivational interviewing. In *CLPsych@ HLT-NAACL*, pages 71–79.
- Vardoulakis, L. P., Ring, L., Barry, B., Sidner, C. L., and Bickmore, T. (2012). Designing relational agents as long term social companions for older adults. In *International Conference on Intelligent Virtual Agents*, pages 289–302. Springer.
- Viera, A. J., Garrett, J. M., et al. (2005). Understanding

⁴Email authors for release details and to obtain the corpus.

interobserver agreement: the kappa statistic. *Fam Med*,
37(5):360–363.