

Towards a Virtual Assistant Health Coach: Corpus Collection and Annotations

Itika Gupta¹ and Barbara Di Eugenio¹ and Brian Ziebart¹ and Bing Liu¹ and Ben Gerber² and Lisa Sharp² and Rafe Davis² and Aiswarya Baiju¹

¹ Department of Computer Science, ² Institute for Health Research and Policy
University of Illinois at Chicago

1 Introduction

Many health problems faced by individuals can be mitigated with changes in health behavior. However, successfully implementing healthy behaviors in ones daily life requires significant motivation that most people, individually, find difficult to initiate and maintain. Health coaching has been identified as a successful method for motivating and maintaining health behavior changes by having a peer or professional convey relevant medical information, help to set realizable, yet challenging goals tied to health behavior change, and provide encouragement in adhering to those goals (Kivelä et al., 2014). But unfortunately, personal health coaching is time-intensive, uneconomical for the low-income patients, and has limited accessibility.

Therefore, we aim to create a dialogue-based virtual assistant health coach that will converse with the patient via text messages and will help them increase physical activity by setting Specific, Measurable, Attainable, Realistic and Time-bound (SMART) goals (Doran, 1981). Even though an influential strand of work on conversational agents has been conducted by Bickmore and his group, their systems rely on pre-defined set of utterances from the patients (Bickmore et al., 2015). We plan to build an autonomous system that learns from observed communications between human health coach and participants to increasingly automate the generation of dialogue (Rizzolatti and Craighero, 2004). The ability to recognize proposed health goals in dialogues is a key initial capability needed for this type of learning from demonstration.

In this paper, we talk about our data collection process, two annotation schemas, agreement results and future work on extracting goals from patient-coach dialogue.

2 Data Collection

We recruited 28 patients between the age of 21 to 65 years who were interested in increasing their physical activity. A health coach, trained in SMART goal setting, conversed with these patients to set goals on weekly basis for a month via SMS. The patients were given Fitbit Alta to monitor their progress. The coach also monitored patients' progress using Fitbit application. The conversation involved setting a specific, measurable and realistic goal, and establishing any

barriers to goal attainment. The coach also sent reminders based on patient's preference and provided motivational feedback on their progress. Only one patient didn't complete the study. We have a corpus of 2857 messages. Among these messages, approximately 53% were sent by the coach and 47% by the patients. This tells us that both the coach and the patients were equally involved in setting or negotiating the goal.

3 Data Annotation

In this section we will look into two types of annotation schemas designed by us to help automate the extraction of health goal set by the user.

3.1 SMART Goal Annotation

3.1.1 Annotation Schema

15 patient-coach conversations were used to design the SMART goal annotation schema shown in Figure 2 (Bovend'Eerd et al., 2009). We didn't annotate Timeliness as a new goal was set every week, and hence by default its value is one week. Each of the annotations can either be categorized as a *slot value* or an *intention*. A slot value is a word or group of words that captures a particular piece of information such as 'walk' is a slot value for *specific activity*. Where as intention is an utterance that tries to gain information about a slot. Each of the SMART annotation category can have other optional tags such as *previous* to annotate an attribute related to previous week's goal, *accomplished* or *remaining* to annotate the progress of the patient, *update* to add another slot value to an existing one, and *other* for anything which doesn't belong to previous or current week. Figure 1 shows the use of SMART Annotation schema.

3.1.2 Annotation Results

Two annotators annotated four previously unseen patients' data for SMART goals. Inter annotator agreement (IAA) was measured using kappa statistics on individual SMART categories and the results are shown in Table 1.

We measured kappa on two levels: message and word. In message level, we consider an agreement if both the annotators labeled at least one word in the

Coach: What goal could you make that would allow you to do more walking?
Patient: Maybe walk (*S.activity*) more in the evening after work (*S.time*).
Coach: Ok sounds good. How many days after work (*S.time*) would you like to walk (*S.activity*)?
M.days_number_intent
Coach: And which days would be best?
M.days_name_intent
Patient: 2 days (*M.days_number*). Thursday (*M.days_name*), maybe Tuesday (*M.days_name_update*)
Coach: Think about how much walking (*S.activity*) you like to do for example 2 block (*M.quantity_distance_other*)
M.quantity_intent
Patient: At least around the block (*M.quantity_distance*) to start.
Coach: On a scale of 1 – 10 with 10 being very sure. How sure are you that you will accomplish your goal?
A.intent
Patient: 5 (*A.score*)

Figure 1: SMART goal annotations for a conversation between the health coach and the patient.

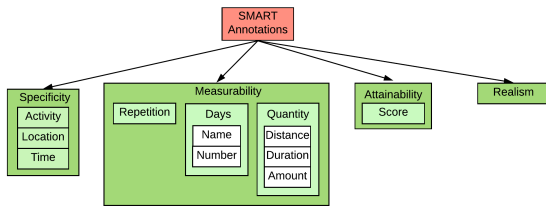


Figure 2: SMART goal annotation schema

Level	S	M	A	R
Message	0.967	0.965	0.907	0.694
Word	0.878	0.895	0.515	0.549

Table 1: Results for Inter-Annotator Agreement.

message with the given tag (not necessarily the same word). In word level, we consider an agreement if both the annotators labeled the same word with the given tag.

In total, 447 messages were annotated for IAA. There were 128 messages with *Specificity* tag, 120 with *Measurability* tag, 45 with *Attainability* tag and 13 with *Realism* tag. We observe approximately 90% reliability for *Specificity* and *Measurability* and only 50% reliability for *Attainability* and *Realism*. This is because {S, M} tags are easy to annotate and have high number of occurrences in the data as compared to {A, R} which are hard to distinguish from each other and have very few occurrences. It should also be noted that for {S, M} word level annotation is more important where as for {A, R} message level annotation makes more sense.

Table 2 shows the statistics for SMART categories over the entire dataset of 2857 messages. One can observe that the percentage of {R} in the entire dataset is

	S	M	A	R
Slot	653	608	60	0
Intent	31	87	260	65
Total	684	695	320	65

Table 2: Messages in each SMART Category.

fairly small when compared to the {S, M, A} tags. It is not surprising as the coach only questions the realism of the goal if he thinks the goal is either too difficult or too easy based on the patient’s past performances.

3.2 Stages and Phases Annotation

3.2.1 Annotation Schema

The aim of this annotation is to understand how the conversation unfolds in a health coaching dialogue. Stages and phases help to capture the coaching tasks and sub tasks being performed throughout the communication dialogue respectively. The higher tier is composed of stages; *goal setting* and *goal action*. Stages are composed of phases. Goal setting stage consists of *goal identification*, *goal refining*, *goal negotiation*, *anticipate barrier* and *solve barrier*. Goal action stage consists of the same phases plus an additional *follow up* phase. A snippet of such conversation is shown in Figure 3. Each message can be annotated as *goal setting*, *goal action*, or none. Phases are optional.

3.2.2 Annotation Results

Similar to SMART annotation, two annotators annotated four previously unseen patients for stages and phases. The confusion matrix showing the annotation counts is shown in Table 3, where ‘gs’ stands for *goal setting*, ‘ga’ stands for *goal action*, ‘I’ stands for *goal identification*, ‘R’ for *goal refining*, ‘A’ for *anticipate barrier*, ‘S’ for *solve barrier*, ‘N’ for *goal negotiation*, and ‘F’ for *follow up*. The kappa statistics for all the categories together came out to be 0.93.

In total, 398 messages were annotated for IAA. Out of them, 115 messages were annotated as *goal setting* whereas 59 messages were annotated as *goal action* by both the annotators. The high value of kappa can be attributed to the fact that while following SMART criteria for goal setting, the stages/phases are bound to occur in a particular order and therefore, are easy to annotate.

Table 4 shows the summary of stages and phases annotations over the entire dataset of 2857 messages. Many categories have zero occurrence in the data used for IAA as they are not very frequent in the entire dataset.

Coach: What goal could you make that would allow you to do more walking? (*Goal Setting - Goal Identification*)
Patient: I could walk on my lunch break for 20 mins Wed and Thurs (*Goal Setting - Goal Identification*)
Coach: Good answer
Coach: What do you think will make it easy to accomplish/achieve your goal? (*Goal Setting - Anticipate Barrier*)
Patient: I just have it set in my mind.... And once I set a goal I like to see it through (*Goal Setting - Anticipate Barrier*)
Coach: I can also send some reminders. What time do you take lunch? (*Goal Setting - Goal Refining*)
Patient: Ok that's sounds good.... Normally around 1030-11 (*Goal Setting - Goal Refining*)
Coach: Sounds like a plan. Good luck!! (*Goal Action*)
Patient: Thanks
Coach: Don't forget to walk on lunch break
Patient: Ok perfect timing....I'm going on break now
Coach: Good morning! How is your goal for this week going so far? (*Goal Action - Follow up*)
Patient: Wed I only walked 10 mins of my break due to the rain. Thur I did the full 20min (*Goal Action - Follow up*)
Coach: You can get another 10 min in today or Monday so you can still accomplish your goal. (*Goal Action - Solve Barrier*)
Patient: Great idea!! I will try to get that in later today (*Goal Action - Solve Barrier*)
Coach: Good morning! Were you able to get that 10 minute walk in last week? (*Goal Action - Follow up*)
Patient: Yes everything worked out great (*Goal Action - Follow up*)

Figure 3: Stages and Phases Annotations for a Conversation between the Health Coach and the Patient.

	gs_I	gs_R	gs_A	gs_S	gs_N	ga_I	ga_R	ga_A	ga_S	ga_N	ga_F	ga	None
gs_I	32	0	0	0	0	0	0	0	0	0	0	0	0
gs_R	1	32	0	0	0	0	4	0	0	0	0	0	0
gs_A	0	0	48	0	0	0	0	0	0	0	0	0	0
gs_S	0	0	1	1	0	0	0	0	0	0	0	0	0
gs_N	0	0	0	0	0	0	0	0	0	0	0	0	0
ga_I	0	0	0	0	0	0	0	0	0	0	0	0	0
ga_R	0	0	0	0	0	0	0	0	0	0	0	0	0
ga_A	0	0	0	0	0	0	0	0	0	0	0	0	0
ga_S	0	0	0	0	0	0	0	0	2	0	0	0	0
ga_N	0	0	0	0	0	0	0	0	0	0	0	0	0
ga_F	0	0	0	0	0	0	0	1	0	0	44	0	3
ga	0	0	0	0	0	0	0	0	0	0	0	12	3
None	0	0	1	0	0	0	0	0	0	0	2	1	210

Table 3: Confusion Matrix for Stages and Phases Annotations.

gs_I	gs_R	gs_A	gs_S	gs_N	ga_I	ga_R	ga_A	ga_S	ga_N	ga_F	ga
252	245	291	25	36	1	15	7	8	12	217	102

Table 4: Messages in each Stage and Phase Category

4 Discussion and Future Work

In this paper we discussed our data collection process, SMART goal annotation schema, and stages and phases annotation schema. We plan to use these annotations to train a classifier that can extract and summarize the goal set by the patient. We are currently working on a pipeline that will first detect all the possible SMART goal features in the conversation, then detect the stages/phases in the conversation and then finally combine these two to extract the correct SMART goal. For example, if a goal was set during the goal identification phase, but then due to some barrier the goal changed in the goal negotiation phase, the summarization model should summarize the updated goal.

References

Timothy Bickmore, Reza Asadi, Aida Ehyaei, Harriet Fell, Lori Henault, Stephen Intille, Lisa Quintiliani,

Ameneh Shamekhi, Ha Trinh, Katherine Waite, et al. 2015. Context-awareness in a persistent hospital companion agent. In *International Conference on Intelligent Virtual Agents*, pages 332–342. Springer.

Thamar JH Bovend'Eerd, Rachel E Botell, and Derick T Wade. 2009. Writing smart rehabilitation goals and achieving goal attainment scaling: a practical guide. *Clinical rehabilitation*, 23(4):352–361.

George T Doran. 1981. Theresa smart way to write managements goals and objectives. *Management review*, 70(11):35–36.

Kirsi Kivelä, Satu Elo, Helvi Kyngäs, and Maria Kääriäinen. 2014. The effects of health coaching on adult patients with chronic diseases: a systematic review. *Patient education and counseling*, 97(2):147–157.

Giacomo Rizzolatti and Laila Craighero. 2004. The mirror-neuron system. *Annu. Rev. Neurosci.*, 27:169–192.