Creating and Annotating a Corpus of Health Coaching Dialogue

Itika Gupta[1] and **Barbara Di Eugenio**[1] and **Brian Ziebart**[1] and **Bing Liu**[1] and **Ben Gerber**[2] and **Lisa Sharp**[2] and **Rafe Davis**[2] and **Aiswarya Baiju**[1]
[1] Department of Computer Science
[2] Institute for Health Research and Policy
University of Illinois at Chicago
Chicago, IL USA
{igupta5, bdieugen, bziebart, liub,
bgerber, sharpl, rdavis5, abaiju2}@uic.edu

Abstract

In this paper, we present our data collection process, annotation schema and agreement results for extracting health goals from SMS conversations between a health coach and the patients. This is our first step towards building an autonomous virtual assistant health coach that learns from expert demonstration to interact with patients via SMS.

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; Wolever et al., 2010). 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 patients 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 predefined set of utterances from the patients (Bickmore et al., 2015, 2011; Schulman and Bickmore, 2009). 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, annotation schema, agreement results and future work on extracting goals from patient-coach dialogue.

2 Annotation Process

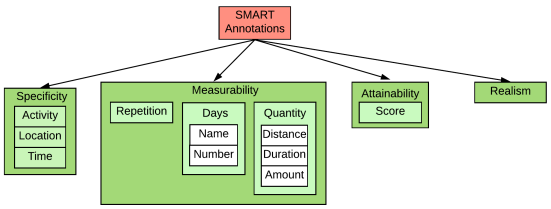

Figure 1: SMART goal annotation schema

In this section we discuss the data collection process and annotation schema for SMART goals.

2.1 Corpus

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 did not complete the study.

We have a corpus of 2858 messages in total, where approximately 54% of messages were sent

Coach: Also, what would you like to set as your SMART goal this week?

[Message 2, 3 omitted for brevity]

Patient: My goal for this week is 2 miles(*M_quantity_distance*) a day(*M_repetition*)

[Message 5, 6 omitted for brevity]

Coach: Okay so you want to stick with your same goal as last week? FYI 2 miles(*M_quantity_distance*) is 4000 steps(*M_quantity_amount*)

Patient: Let's make it 3 miles(*M_quantity_distance*)

Coach: Base on your steps(*S_activity*) the first week I think 3 miles(*M_quantity_distance*) a day(*M_repetition*) is doable... (R_intent)

Coach: *How many days do you want to go for?* (M_days_number_intent)

Patient: Five(*M_days_number*)?

Patient: I think I can do it all my days

Patient: Let's go for a week(*M_days*)

Coach: ok so all 7 day(*M_days_number*)?

Patient: Yes

Coach: *What do you think will make it easy to accomplish/achieve your goal?* (A_intent)

Coach: *Me walking*(*S_activity*) *everyday*(*M_days*) (A_intent)

Figure 2: SMART goal annotation for a conversation between the health coach and the patient.

Level	Specificity	Measurability	Attainability	Realism
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.

by the coach and 46% by the patients. This tells us that patients were equally involved in setting the goal. We plan to collect more data with different health coaches in the next phase of the project. Unfortunately, the dataset cannot be released because of human subject protection restrictions.

2.2 SMART Annotation Schema

15 patient-coach conversations were used to design the annotation schema (Bovend'Eerdt et al., 2009). After four iterations, we finalized the schema for annotating SMART goals as shown in Figure 1. We didn't annotate Timeliness as a new goal is set every week, and hence by default it 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* for annotating anything related to previous week, *accomplished* or *remaining* is used 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 2 shows the use of SMART annotation schema for a small sample of conversation from our dataset.

3 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. (Cohen, 1960)

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 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. There were 128 messages with *Specificity* tag, 120 with

	Specificity	Measurability	Attainability	Realism
Slot	**653**	**608**	60	0
Intent	31	87	**260**	**65**
Total	684	695	320	65
Total (%)	23.93	24.32	11.20	2.27

Table 2: Messages in each SMART category.

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* on word level. 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 2858 messages. One can observe that the percentage of {R} in the entire dataset is 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.

4 Discussion and Future Work

In this paper we discussed our data collection process and SMART goal annotation schema. We plan to use these annotations to train a classifier that can extract and summarize the SMART goal set by the patient. Using Stanford Temporal Tagger (SUTime) and rules such as Part-Of-Speech followed by a given keyword, we are able to obtain a minimum F1 score of 70% on all the SMART categories. We think that detecting different goal setting stages such as *goal identification*, *goal negotiation*, *anticipate barriers*, and *follow-up* can help us to improve the results further. Therefore, we are working on annotating the data for stages and will build a classifier for the same. Our aim is to combine these two classifiers to present the summary of the goal set by the patient. This will be the first step towards building a health coaching dialogue system for increasing physical activity.

Acknowledgments

This work is supported by NSF grant, EAGER: The Virtual Assistant Health Coach: Summarization and Assessment of Goal-Setting Dialogues. Award Number:1650900

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.

Timothy W Bickmore, Daniel Schulman, and Candace L Sidner. 2011. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. *Journal of biomedical informatics*, 44(2):183–197.

Thamar JH Bovend'Eerdt, 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.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

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.

Daniel Schulman and Timothy Bickmore. 2009. Persuading users through counseling dialogue with a conversational agent. In *Proceedings of the 4th international conference on persuasive technology*, page 25. ACM.

RQ Wolever, M Dreusicke, J Fikkan, TV Hawkins, S Yeung, J Wakefield, L Duda, P Flowers, C Cook, and E Skinner. 2010. Integrative health coaching for

patients with type 2 diabetes a randomized clinical trial. *The Diabetes Educator*, 36(4):629–639.