Predicting Client’s inclination towards Target Behavior Change in Motivational Interviewing and investigating the role of laughter

Rahul Gupta¹, Panayiotis G. Georgiou¹, David C. Atkins² and Shrikanth Narayanan¹

¹Signal Analysis and Interpretation Lab (SAIL), Department of Electrical Engineering
University of Southern California, Los Angeles, CA
²Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA

Abstract

Motivational interviewing (MI) is a goal oriented psychotherapy that facilitates intrinsic motivation within a client in order to change behavior in a dialog setting. The Motivational Interviewing Skills Code (MISC) is a manual observational coding method used to quantify and evaluate the quality of MI sessions using their audio-visual recordings. However, this coding method is both labor intensive and expensive. We present an approach towards automating MISC assignments in MI involving addiction cure. Specifically, we focus on predicting valence for “Client Change Talk” (ChangeTalk) utterances, which indicate a client’s attitude towards a “Target Behavior Change” (Target). We further study the effect of incorporating counselor behavior in the model. We observe that our best model achieves an unweighted accuracy of 50.8% in a 3-way classification of positive vs negative valence ChangeTalk vs no ChangeTalk. Furthermore, we study the effect of including non-verbal behavior, specifically laughter, in our model. Information regarding location of laughter improves the unweighted accuracy of our model to 51.4% and our experimental results suggest prosodic differences in laughter belonging to ChangeTalk utterances with different valences.

Index Terms: Motivational Interviewing (MI), Motivational Interviewing Skill Code (MISC), Client Change Talk (ChangeTalk) utterance, Maximum entropy model.

1. Introduction

Rollnick et al. [1] define Motivational Interviewing (MI) as “a directive, client-centered counseling style for eliciting behavior change by helping clients to explore and resolve ambivalence”. MI setting is extensively used in addiction related problems (alcohol, drugs, etc.) [2, 3]. MI helps addiction patients perceive both, the benefit (e.g., the high) and the harm (e.g., health, personal problems) and helps them resolve the ambivalence in a dyadic spoken interaction with a therapist towards a positive change. With emerging evidence base and popularity for MI [4], a significant challenge lies in ensuring high quality of treatment which calls for standard metric to assess quality of such interactions. The Motivational Interviewing Skill Code (MISC) [5] has emerged as a standard observational method for evaluating the quality of MI interactions. MISC is a behavioral coding systems in which human coders are trained to annotate video or audio tapes over several parameters as global counselor ratings, empathy, behavior categories etc.. The uses of MISC range from providing detailed session feedback to counselors in the process of learning MI to predicting treatment outcome from psychotherapy measures.

MISC provides us with a systematized tool to assess the quality of MI interactions. However such manual coding systems are not scalable to real world use due to time, labor and economic constraints [6]. As part of our ongoing efforts towards Behavioral Signal Processing [7, 8], in this paper, we present a model towards the automation of MISC annotation. Specifically, we focus on the “Target Behavior Change” (Target) aspect of the client behavior that specifies a target client behavior (smoking, drinking, medication) and a direction of change (stopping, increasing etc.). We conduct three sets of experiments to (a) predict a client’s attitude towards a Target, (b) investigate the effect of laughter during interaction on Target and (c) look for prosodic differences in laughter with respect to different behaviors.

(a) Predicting client’s attitude towards Target: A turn in client speech is annotated with a positive/negative valence if it shows an inclination towards/away from Target. Such client turns are termed as “Client Change Talk” (ChangeTalk) utterances. We design a lexical based model for predicting positive vs negative valence ChangeTalk vs no ChangeTalk given a client utterance. Additionally, we study the effect of including counselor behavior in our prediction model. Each counselor utterance is MISC annotated with a behavior code (reflect, support etc.). As the counselors are required to carefully attend to client language related to the target behavior, their behavior may carry indicators to the client’s attitude towards Target. Our best prediction model achieves an unweighted accuracy (UA) of 50.8% (chance 33.3%) for the three way classification.

(b) Investigating the role of laughter during interaction: Laughters are linked to human behavior and are hypothesized to carry out several social functions [9, 10]. In this experiment, we examine if the mere occurrence of a laughter during interaction carries some information with regards to the ChangeTalk valence. We observe that inclusion of information regarding their presence improves UA of our previous model to 51.4%.

(c) Prosodic differences amongst laughters: Several studies [11, 12, 13] suggest differences amongst laughters contingent upon the context in which they happen. We look for prosodic differences amongst client laughters belonging to utterances in the three ChangeTalk classes. We observe some discriminatory power in their prosody and achieve an UA of 40.5% in classifying laughters belonging to utterances from the three classes.

2. Database

Our experiments pertain to an MI based intervention study on drug abuse problems involving patients at a public, safety-net hospital [14, 21, 22]. We use data from 49 subjects coded...
3. Experiments

3.1. Predicting valence of Client Change Talk utterance.

Although modeling the MISC annotation of an MI session may be very complex, in this work we propose a simplified scheme to address this problem. We represent each counselor utterance in utterance (9060) and counselor behavior codes in Table 1. Utterances with positive valence during ChangeTalk are listed as Change+, negative valence as Change- and with no change talk as Change0. Table 2 lists the count of utterances belonging to various ChangeTalk and counselor behavior codes.

### Table 1: Example conversation with Counselor (T) behavior annotation and client (C) ChangeTalk valence annotation.

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Counselor behavior code</th>
<th>Change Talk</th>
</tr>
</thead>
<tbody>
<tr>
<td>T: Did you come here on own?</td>
<td>Question</td>
<td>Change+</td>
</tr>
<tr>
<td>C: Yes, I was sure about this.</td>
<td>Support</td>
<td>Change0</td>
</tr>
<tr>
<td>T: I am here to help you.</td>
<td></td>
<td>Change-</td>
</tr>
<tr>
<td>C: I appreciate that.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C: I really have to stop drugs,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C: but I just don’t want to.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: List of various counselor behavior codes and ChangeTalk codes and corresponding count of utterances.

<table>
<thead>
<tr>
<th>Client change talk utterance</th>
<th>Positive valence (1749)</th>
<th>Negative valence (1253)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No ChangeTalk in utterance (9060)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Figure 1: Proposed model to represent the dyadic interaction and the annotation process.

This assumption. In this baseline model, we focus on lexical content of client utterances and design an n-gram based classification system. We describe our feature extraction, selection and classification scheme below. We perform a leave-one-session-out cross validation in all our experiments.

(1) Feature extraction: Given the unbalanced class distribution, we initially downsample instances from the majority classes (Change0, Change+), so that each class has the same number of instances as the least represented class; Change-. We extract all the unigrams and bigrams from each client utterances as potential features to learn the corresponding ChangeTalk code. However, this leads to a large feature space with several features that may not be relevant to the classification or may contain n-grams that are rare in occurrence. To overcome this problem, we use a feature selection algorithm as described next.

(2) Feature selection: We select a given n-gram \((n_{g})\) based on the entropy \((E(n_{g}))\), equation (1) of its empirical distribution (equation 2) over the three classes and their minimum count \((#n_{g})\) on the downsampled data as shown in the Algorithm 1. The maximum entropy threshold \((T_{E})\) and the minimum word count \((C_{min})\) thresholds in the algorithm are tuned on a subset of the training set.

#### Algorithm 1: N-gram selection for classifier training.

1. Define: \(G = \{n_{g1}, ..., n_{gK}\}\) : Set of n-grams in the training set.
2. \(G_{sel}\): Set of selected n-grams for classification
3. Initialize: \(G_{sel} = \emptyset\)
4. for \(k = 1 \ldots K\) do
5.   if \((E(n_{g}) < T_{E})\) and \((#n_{g} > C_{min})\) then
6.     \(G_{sel} = G_{sel} \cup n_{g}\)
7.   end if
8. end for

\[
E(n_{g}) = \sum_{\text{ChangeTalk}_{cd} \in \{\text{Change+}, \text{Change-}, \text{Change0}\}} \log (P_{k_{cd}}) \times \frac{C_{cd}}{\#n_{g}}
\]

where, \(P_{k_{cd}} = \frac{C_{cd}}{\#n_{g}}\) in Training utterances coded as ChangeTalk

#### Classifier training: We train a conditional maximum entropy model [15] on the chosen set of n-grams \((G_{sel})\). We perform the parameter estimation using L-BFGS [16]. Our baseline system can be viewed as only utilizing the solid line connecting \(U_{Cm}\) to \(C_{cm}\) in Figure 1, i.e.: ChangeTalk code depends only on what client says. Equation 3 shows the decision
rule for assigning class the ChangeTalk class $C_m$ to the utterance $U_m$, given a set of observed features $O_m$. In our baseline model, we set $O_m \rightarrow G_m, O_{m+1}$ (excluding $G_m$), the set of $n$-grams extracted from the considered utterance $U_m$.

$$C_m = \arg \max_{\text{ChangeTalk}_m} P(\text{ChangeTalk}_m \mid O_m)$$

(3)

### 3.1.2. Classification system incorporating counselor behavior

In our next experiment, we account for the counselor behavioral codes while inferring the ChangeTalk codes, i.e. the client’s utterance and the counselor’s immediate past behavior both contribute towards identifying the code. Hence, in this scheme we also utilize the dotted link connecting the preceding counselor behavior code $B_{T_c \leq m-1}$ in addition to features from $U_m$ to infer $C_m$. We perform two sets of experiments incorporating oracle and inferred counselor behavior codes as follows.

**Oracle counselor behavior:** In this experiment, we use the oracle counselor behavior code ($B_{T_c \leq m-1}^o$) preceding $C_m$ as annotated by the coder. This model utilizes the list of features $\{G_{m}, B_{T_c \leq m-1}^o\}$ as the evidence in equation 3. Note this is not ideally possible in a real system as we are using the true counselor codes for the test set which likely won’t be available in the real world scenario.

**Inferred counselor behavior:** As the use of oracle values for oracle counselor behavior is impractical, we develop a system to infer them. We implement the same framework as described in the baseline system. However as we have extremely low number of training instances for a few codes, we merge the few minority classes before training our prediction system. We retain questions (QU), giving information (GI), reflection (RE) and facilitate (FA) and merge all the other classes into a fifth class; others (OT). Initially, we gauge the effect of merging counselor codes on our previous prediction system involving oracle counselor codes. $O_m$ in equation 3 is set to $\{G_{m}, B_{T_c \leq m-1}^o\}$, where $B_{T_c \leq m-1}^o$ are the oracle counselor codes obtained after merging.

In order to predict the merged counselor codes, we initially downsample the data to remove class bias over the 5 classes. We perform feature selection and classifier training as described in the baseline system for inferring the counselor codes using the n-grams from counselor utterances ($U_m$). Equation 4 shows the rule for inferring the counselor code $B_{T_c \leq m-1}^{\text{pred}}$ from the set of features $O_{T_c \leq m-1}$. In this model, $O_{T_c \leq m-1}$ is set to the selected set of n-grams $G_{m}, B_{T_c \leq m-1}^{\text{pred}}$ extracted from counselor utterance $U_T$, preceding $U_m$. We use the inferred counselor code as the observed evidence $O_m$ in equation 3 ($O_m = \{G_{m}, B_{T_c \leq m-1}^{\text{pred}}\}$).

$$B_{T_c \leq m-1}^{\text{pred}} = \arg \max_{\text{pred}} P(\text{pred} \mid G_{m}, B_{T_c \leq m-1}^{\text{pred}})$$

(4)

### 3.1.3. Results and discussion

We use unweighted accuracy (UA) as our evaluation metric and also report the F-measure for Change+ and Change- given their low occurrence frequency relative to Change0. The results for inferring ChangeTalk codes using the baseline system and after incorporating $B_{T_c \leq m-1}^{\text{pred}}$ are shown in Table 3. We also show the results for inferring counselor behavior codes in Table 4.

**Discussion:** From the results we observe that our model performs well above chance, indicating that just lexical content during conversation can inform us of change talk behavior. Improvement in the results, after including the counselor behavior code indicates, that coder does take context of conversation into account while assigning the ChangeTalk codes. Also, we do not observe any significant difference between using all the behavior codes versus using merged oracle codes. This suggests that training the model on a few codes with sufficient number of samples is as good as training on all the codes with a few samples. However, we do observe a decrease when we use the inferred counselor codes over oracle codes. This stems from the imperfect prediction of the counselor codes themselves. We observe that a few classes in case of counselor codes are predicted more accurately as compared to others in spite of using balanced number of instances during training. This indicates that some classes are better distinguishable with lexical features, while other classes may be more distinguishable in other modalities. Particularly in the case of facilitate (FA), as most of utterances are simple, functioning as “keep going” acknowledgment such as “Mm Hmmm”, “OK” etc., we observe almost perfect prediction.

### 3.2. Prediction incorporating laughters

Several studies suggest the importance of laughters in discourse [9, 10]. In this section, we perform preliminary analysis of laughters and study their effect on our previous system. We hypothesize that mere occurrence of laughter events may provide us with some information regarding a client’s attitude regarding Target. We add a simple binary feature indicating presence of laughter $L_{T_c}$ (available through transcripts) in the client utterance $U_m$, to our previous models and reproduce the results. We set $O_m$ to $\{G_{m}, L_{T_c}\}$ for the baseline model and the similar addition is made to models incorporating the counselor behavior. While inferring $B_{T_c \leq m-1}$, we use a binary features $L_{T_c}$ indicating the presence of the counselor laughter in $U_T$, $O_{T_c} = \{G_{m}, L_{T_c}\}$ in equation 4. We show the results for predicting ChangeTalk codes using laughters and relative improvements over the counterparts from the previous model in Table 5. Results and corresponding relative improvements for counselor behavior code prediction are shown in Table 6.

**Discussion:** From the results, we observe that we get a consistent gain in both the client and the counselor results. We list the empirical distribution of client laughters over the three ChangeTalk codes in Figure 2(a) and the counselor laughters over the merged counselor behavior codes in Figure 2(b). The occurrence probability of laughters is not uniform over codes, thereby providing additional information to the previous model. We observe that in the case of ChangeTalk codes, an utterance labeled Change+ is most likely to contain laughters. This indi-
Table 5: Results (%) & relative improvements (%) over previous model for predicting ChangeTalk codes w/ laughter.

<table>
<thead>
<tr>
<th>Model</th>
<th>Change+ F-meas</th>
<th>Acc. /Prec.</th>
<th>Change- F-meas</th>
<th>Acc. /Prec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(U_{cn} )</td>
<td>50.1 (2.2)</td>
<td>32.1 (5.9)</td>
<td>45.8/24.7 (2.2/7.7)</td>
<td>28.6 (0.0)</td>
</tr>
<tr>
<td>(+B^{T &lt; c_{&lt; m - 1 &gt;}})</td>
<td>51.4 (1.2)</td>
<td>33.4 (2.1)</td>
<td>45.4/26.4 (1.3/2.3)</td>
<td>29.6 (1.7)</td>
</tr>
<tr>
<td>(+B^{T &lt; cc_{&lt; m - 1 &gt;}})</td>
<td>51.3 (1.4)</td>
<td>33.4 (2.5)</td>
<td>45.3/26.4 (1.3/3.1)</td>
<td>29.5 (1.4)</td>
</tr>
<tr>
<td>(+B^{T &lt; m_{&lt; m - 1 &gt;}})</td>
<td>50.8 (1.2)</td>
<td>32.8 (3.1)</td>
<td>44.1/26.1 (1.1/4.4)</td>
<td>29.3 (0.6)</td>
</tr>
</tbody>
</table>

Table 6: Results(%) & relative improvements (%) over previous model for predicting counselor behavior codes w/ laughter.

<table>
<thead>
<tr>
<th>Model</th>
<th>RE</th>
<th>GI</th>
<th>QU</th>
<th>FA</th>
<th>OT</th>
</tr>
</thead>
<tbody>
<tr>
<td>(U_{ct} )</td>
<td>71.3 (2.4)</td>
<td>71.4 (1.0)</td>
<td>56.0 (0.5)</td>
<td>77.2 (0.0)</td>
<td>97.0 (-0.1)</td>
</tr>
</tbody>
</table>

3.3. Laughters and their prosodic differences

Several studies show that there are inherent differences in laughters and the context in which they occur [11, 12, 13]. We investigate the differences amongst client laughters in the context of the three ChangeTalk classes. We design a prosody based classification system to distinguish amongst laughters that occur in utterances coded as Change+, Change- or Change0. We hypothesize that if indeed differences exist amongst the laughters, this can further aid our ChangeTalk valence prediction system. We describe the prosodic features and the classification system below.

**Features:** We manually annotate all the 371 client laughters (Change+:61, Change-:26, Change0: 284) in the 49 sessions marking their start and end positions. Given a small number of samples, we limit our experiment to a few low level prosodic cues and compute their global statistics shown in Table 7. We mean normalize these features per speaker.

**Classifier:** We use a linear kernel support vector machine classifier. Given the unbalanced class distribution (counts shown in Figure 2), we downsample the samples from Change0 and Change+ classes so that each class has equal number of samples. We perform leave one session out cross-validation on the laughters from 49 sessions. We list the classification accuracy in Table 8.

**Discussion:** We observe that the use of a few low level prosodic cues does provide us with some discriminatory power in between client laughters belonging to the three classes. Poor classification accuracy stems from extremely small number of training instances. Laughters from the class Change0 are most poorly classified as they have a high downsampling factor for maintaining the class balance. This introduces a sampling bias. Due to the same reason, we could not carry out an experiment to investigate differences in counselor laughters as some classes have too few samples (e.g. 7 for QU). Because of the weak discrimination, this information does not help our previous ChangeTalk valence prediction model as of now. However, given that prosodic differences in laughters do exist, we are encouraged that we improve the proposed model with the availability of more training data in future.

4. Conclusion

In this paper, we present a scheme towards automatically obtaining MISC codes in MI settings. We design a lexical based scheme to automatically identify client utterances with positive or negative valence that indicate their attitude towards a targeted behavior change. We show that incorporating the counselor’s behavior into account during the interaction helps improve the prediction, thus validating the importance of the counselor towards positive outcomes. We proceed with a preliminary analysis on incorporating laughters as additional information source, augmenting the previous system. Finally, we analyze the type of laughters based on their prosodic cues. We observe some discriminatory power in the prosody of laughters with respect to the ChangeTalk valences, however due to limited data and poor accuracy we could not exploit this towards change talk classification.

We presented our results on one aspect of MISC code. However, the MISC annotations provide other global measures such as empathy, motivational interviewing spirit etc. that furnish more indicators regarding the success of a session. We aim at building upon our current system to incorporate these measures. Studies link acoustic [17, 18], visual cues [19, 20] etc. to human behavior and one can incorporate such cues to supplement the system. Also, we aim to further investigate other aspects of laughters (e.g. shared laughters) and the role they play in MI. One may further extend this work to other non-verbal vocalizations as fillers, sighs etc.

5. Acknowledgement

This research is supported by NSF, NIH and DARPA.
6. References


