Predicting Success of Clinical Interviews via Probabilistic Modeling of Patient-Provider Communication Sequences

Mehedi Hasan, BS[1], Alexander Kotov, PhD[1], April Carcone, PhD[2], Mind Dong, PhD[1], Sylvie Naar, PhD[2]
[1]**Department of Computer Science,** [2]**Pediatric Prevention Research Center, Wayne State University, Detroit, Michigan**

Abstract *The problem of analyzing temporally ordered sequences of past events to make predictions related to the observed sequence of events arises in many data analytics scenarios in healthcare informatics. In this paper, we focus on automating the analysis of patient-provider communication sequences in the context of clinical interviews and propose the methods for identifying the most effective communication exchanges to elicit a particular type of patient response and dynamically predicting this type of response in an observed patient-provider communication. Our proposed method for predicting the likelihood of desired patient behavioral response in communication sequences can be used in conjunction with either Markov Chain or Hidden Markov Model and can significantly decrease the effort required to develop effective interventions to address many public health conditions, such as childhood obesity. It achieved 75.32% and 79.89% F-measure for natural sequences and 97.36% and 96.99% F-measure for alternating sequences using first and second order Markov Chain and Hidden Markov Model, respectively, to model communication sequences. Our results indicate that individual short length patient-provider communication sequences are highly indicative of the overall progression and future trajectory of a clinical interview and can be used to predict its overall success. Our results have broad implications for qualitative public health research by providing a formal mechanism to facilitate the development of effective behavioral interventions.*

Introduction

Data in the form of temporally ordered sequences of discrete or continuous observations (e.g., symbolic sequences such as notes in patient EHR, diagnostic codes, protein or DNA sequences or continuous time series, such as ECG measurements) arise in various domains of health informatics. The order of observations in a sequence is challenging to capture using features, which makes sequence classification a more challenging task than traditional classification. In general, sequence classification methods can be divided into three categories: feature-based, distance metric-based and model-based method. Feature-based classification methods transform a sequence into a feature vector and apply a standard supervised learning algorithm, such as support vector machine[1] or decision tree[2]. Shapelet[3] and pattern[4,5] based techniques as well as hierarchical approaches[6] have been proposed instead of standard classifiers as well. Distance-based methods measure the similarity between sequences to determine the quality of classification. The most commonly used distance function is Euclidian distance[7] with Dynamic Time Wrapping[8] used for more flexible matching in time series data. The third type of sequence classification methods, first create a probabilistic model of a sequence, such as Hidden Markov Model[9] (HMM).

Sequence classification has a broad range of real word applications from genomics and health informatics to finance and anomaly detection. In genomics research, sequence classification is widely used to classify protein and text sequence data[10] and detect the function of new proteins[11]. In health informatics, ECG measurements are considered as multi-dimensional time series and are used to classify individuals as healthy or having a heart disease[12]. Sequence classification is also used for anomaly detection such as abnormal access to systems[13] and malware[14].

In this paper, we direct our focus towards patient-provider communication sequences in the context of clinical interviews. Specifically, we focus on the transcripts of Motivational Interviews (MI) with obese adolescents and their caregivers. Childhood obesity is a serious public health concern in the United States and worldwide. Recent estimates indicate that approximately one third (31.8%) of US children age 2-19 years are overweight and 16.9% are obese[15]. Adolescents who are obese likely continue to be obese in adulthood and have a greater risk of heart disease, type 2 diabetes, stroke, cancer, and osteoarthritis[16]. One approach to designing an effective obesity intervention is Motivational Interviewing, an evidence-based counseling technique to increase intrinsic motivation and self-efficacy for health-related behavior change. The goal of MI counseling session is to encourage patients to explore their own desires, ability, reasons, need for and commitment to the targeted behavior change. These statements, referred to as "change

talk" (or CT), consistently predict actual behavior change that can be sustained for as long as 34 months after an interview. However, the ability of counselors to consistently elicit this type of patient communication requires knowledge of effective communication strategies for a variety of patients, which can only be obtained through analysis of a large number of annotated interviews. Since manual annotation and examination of transcripts is a very time-consuming process, development of new obesity interventions can take years. Therefore, there is a need for informatics-based methods to facilitate the development of effective MI-based interventions.

While our previous work[17, 18] explored several machine learning methods for automatic annotation of clinical interview fragments with a specialized codebook containing 115 patient and provider behavior codes[19], in the work, we propose probabilistic methods to identify patient-provider communication sequences that are likely to elicit the desired patient behavioral response (i.e. change talk or commitment language) and to dynamically estimate the likelihood of observing this desired response at any point during a clinical interview based on all coded previous patient-provider communication exchanges in the same interview.

While sequential analysis[20] was performed on physician-patient dialogue to establish the relationship between physician and patient behaviors associated with patients' exhibiting cues, there have been only a handful of previous studies on modeling annotated sequences of patient-provider communication exchanges and predicting a desired patient behavior in the context of clinical interviews.

Methods

Data collection

The experimental dataset for this work was constructed from the transcripts of 37 motivational interviews, which include a total of 11,353 segmented and annotated utterances. Each motivational interview consists of counselor-adolescent and counselor-caregiver sessions. Since the ultimate goal of a motivational interview is to increase the desire and ability of adolescents for a targeted behavior change, we considered only communication sequences from counselor-adolescent sessions and disregarded all communication sequences from counselor-caregiver sessions. The utterances were annotated based on MYSCOPE codebook[19] with 115 distinct behavior codes, which are grouped into the youth, caregiver, and counselor code groups. Utterances were divided into successful and unsuccessful communication sequences. Successful communication sequences result in either positive change talk or commitment language statements by an adolescent, while unsuccessful sequences are the ones that result in negative change talk, negative commitment language or the ones in which neither change talk nor commitment language statements occur. Out of 1072 observed sequences, 880 were positive and 192 were negative. alternating sequences are the possible combination of a natural sequence where two consecutive utterances cannot belong to the same speaker. For example, there are 6 possible alternating sequences (C1, A1, C2, A3, C1, A1, C3, A3, C1, A1, C4, A3, C1, A2, C2, A3, C1, A2, C3, A3 and C1, A2, C4, A3) for the natural sequence C1, A1, A2, C2, C3, C4, A3, where adolescent and counselor codes are represented by A and C, respectively. A total of 18753 alternating sequences have been observed, out of which 18139 were successful and 614 were unsuccessful ones. For each type of sequence (natural and alternating), one probabilistic model was trained using successful sequences and one model was train using unsuccessful sequences. Statistics of experimental dataset are presented in Table 1 and a fragment of an adolescent session transcript is presented in Table 2.

Table 1: Statistics of experimental dataset.

Sequence type	Total sequences	# Successful sequences	# Unsuccessful sequences	Avg. length
natural sequences	1072	880 (82.09%)	192 (17.91%)	5.83
alternating sequences	18753	18139 (96.72%)	614 (3.28%)	12.89

Annotation column in Table 2 shows the sequence of behavior codes from top to bottom, where counselor starts with an open-ended question and gets positive feedback at the end.

Table 2: Fragment of the annotated transcript of a dialogue between a counselor and an adolescent.

Annotation	Description	Speaker	Text
331	Open-ended question, elicit change talk positive	Counselor	do you feel like making healthier choices for your snacks and your meals is something you would be able to do? mm-hmm meaning is that food available for you?
117	Low Uptake, positive	Adolescent	Yes
301	Structure Session	Counselor	okay and thats an important thing for us to think about cause i would not want to help you come up with a plan that you would not be able to do without somebody else help so the last part of your plan is how somebody could be supportive to you meaning how they can help you be successful and so we should choose somebody who you feel like is around often enough
112	Change Talk positive	Adolescent	my um aunt
301	Structure Session	Counselor	okay so lets stick something my aunt can do
112	Change Talk positive	Adolescent	she could when i am doing when i am eating something that i should i could not be eating but so i can choose something healthy she could tell me not to eat it
309	Affirm, low	Counselor	okay that sounds like a really great suggestion

Sequence classification techniques

Generally, a sequence is a temporally ordered set of events. In our study, an event is a behavior code represented as a symbolic value, such as 117 (low uptake, positive), 331 (open-ended question), etc. Given a sequence of behavior codes represented as $S_i = \{x_1, x_2, ..., x_n\}$ and a set of class labels $L = \{l_1, l_2, ..., l_m\}$, the task of sequence classification is to learn a sequence classifier C, which is a function mapping a sequence S_i to a class label $l_i \in L$, written as, $C : S_i \rightarrow l_i, l_i \in L$, where L is fixed and known in advance. In our case, the labels are "successful" and "unsuccessful" (motivational interview).

We model successful and unsuccessful patient-provider interactions using first- and second-order Markov Chain (MC) and Hidden Markov Model (HMM), which are probabilistic methods for modeling discrete observation sequences with finite vocabulary. Although HMM was originally developed for speech recognition[9], it is one of the most widely used methods for sequence modeling[21–25].

Markov Chain: In probability theory, a Markov model is a stochastic model used to model randomly changing systems where it is assumed that future states depend only on the present state and not on the sequence of events that preceded it (that is, it assumes the Markov property). Generally, this assumption enables reasoning and computation with the model that would otherwise be intractable. For the sequential analysis, we built two Markov models M and $\overline{M}$ describing provider strategies and patient responses in case of successful (M) and unsuccessful ($\overline{M}$) motivational interviews. A Markov model M can be represented as a weighted directed graph $G = (V, E, p)$, in which:

- $V = \{CML+, CHT+, CHT-, T-AMB, CCT, BLT, LUP+, LUP-, HUP-W, ...\}$ is a set of vertices, consisting of adolescent and counselor MI behavior codes;

- $E \subseteq V \times V$ is a set of edges corresponding to posssible transitions from one MI behavior code to the other in a sequence;

- $p_M : E \rightarrow [0...1]$ is a function that assigns probability $p(c_i|c_j)$ to an edge between the MI behavior codes c_i and c_j based on maximum likelihood estimator:

$$P_M(c_j|c_i) = \frac{n_{c_i,c_j}}{n_{c_i}} \tag{1}$$

where n_{c_i,c_j} and n_{c_i} are the number of times a transition between the MI behavior codes c_i and c_j and the code c_i has been observed in the training data. Given a Markov model M (such that $S \subseteq V$), the probability that a sequence of MI behavior codes $S = \{C_1, ..., C_N\}$ has been generated from a Markov model M is:

$$P_M(S) = \prod_{i=2}^{N} p_M(c_i|c_1, \ldots, c_{i-1}) = \prod_{i=2}^{N} p_M(c_i|c_{i-1}) \tag{2}$$

The likelihood of success of a given motivational interview at a given time point, given a sequence of MI behavior codes observed prior to that point, is estimated based on the following formula:

$$p(S \to CML+) = \log\left(\frac{P_M(S)}{P_{\overline{M}}(S)}\right) = \sum_{i=2}^{N} \log p_M(c_i|c_{i-1}) - \sum_{i=2}^{N} \log p_{\overline{M}}(c_i|c_{i-1}) \tag{3}$$

if $p(S \to CML+) > \delta$, where δ is an empirically determined threshold, the interview is predicted to result in positive change talk or commitment language. Estimating the model on the training data involves experimentally determining the value of δ, resulting in the highest prediction F-measure with lower bias. Figure 1 illustrates the accuracy of Markov model-based method by varying the value of δ. As can be seen from Figure 1, F-measure varies insignificantly, when δ is between -5 to -1, due to possible bias. However, F-score sharply decreases after $\delta = -0.9$. Thus, -0.9 is the optimal value of δ, where F-measure is the highest with lower bias.
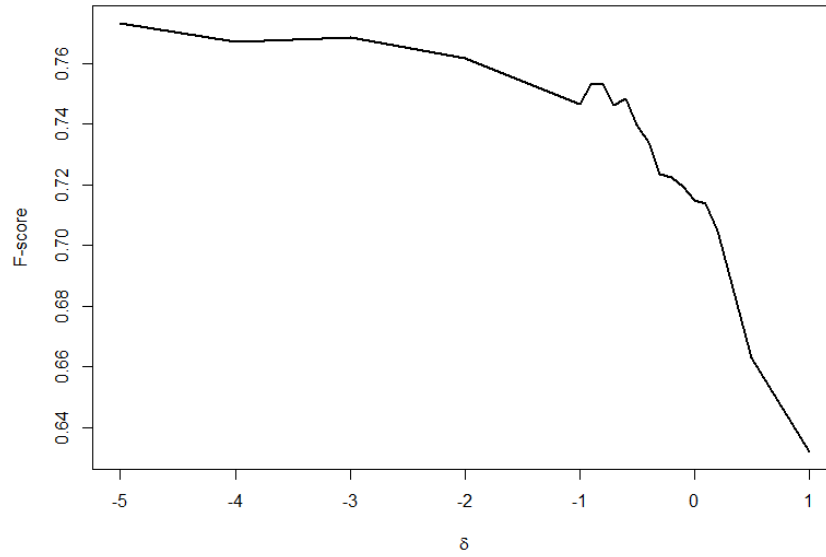

Figure 1: Accuracy of the proposed Markov model-based method by varying the value of δ.

The Markov model we have discussed so far is referred to as the first-order MC, since it only considers immediately preceding behavior code to compute the state transition probability matrix. A generalization of the first-order MC is the kth order MC, in which the transition probability to a particular state (symbol) x_i is computed by looking at the k preceding states (symbols). Thus, k^{th} order Markov chain will have N^k states each associated with a sequence of k symbols. In our experiment, we also used second order Markov model. Hence, our model has N^2 states with transition probability matrix of size $N^2 \times N$.

Hidden Markov Model (HMM) is a powerful tool for statistical modeling of processes varying in time. HMMs are

widely used for sequence analysis because of their ability to incorporate dependencies among elements in sequence. HMM can be considered as a doubly embedded stochastic process with a process that is not observable (hidden process) and can only be observed through another stochastic process (observable process) that produces a sequence of observations. An HMM can be fully specified by the following quintuple: $\lambda = (N, M, A, B, \pi)$

- N is a number of hidden states in the model

- M is a number of distinct observations symbols per state, i.e. the discrete vocabulary size

- A is an $N \times N$ state transition probability distribution matrix $A = \{a_{ij}\}$

- B is an $N \times M$ matrix $B = \{b_j(k)\}$ with observation symbol probability distribution for each state

- π is the initial state distribution vector $\pi = \{\pi_i\}$

We exclude the structure parameters M and N to designate HMM using a compact notation: $\lambda = (A, B, \pi)$. The key difference between HMM and MC is that HMM requires specifying the number of hidden states as a model parameter and then the model figures out a sequence of hidden states that best explains the observations along with state transition probabilities and distributions of observation symbols for each state. The Baum-Welch algorithm was used to estimate the parameters of HMMs for successful and unsuccessful interviews using the corresponding training set, while the Viterbi algorithm was used to determine the most likely sequence of hidden states for a given sequence of observations. After assignment of hidden states, the log-likelihood of successful outcome can be estimated using Eq. 3. As in the case of MC, we also experimented with second-order HMM.

Evaluation methods

The performance of classifiers is reported on the basis of 5-fold cross-validation by using weighted macro-averaging over folds. Precision, recall, F-measure and accuracy were used to evaluate the performance of classifiers. A true positive (TP) was counted when classifier correctly classified the sequence into actual class; a false positive (FP) was counted when classifier incorrectly classified other classes to the considered class; a false negative (FN) when the considered class incorrectly classified to other classes. The precision of a class was defined as the ratio between the numbers of correctly classified sequences and all sequences identified as that particular class by the classifier. In other words, Precision = TP / (TP + FP). Recall of a class was defined as the ratio between the numbers of correctly classified sequences and all sequences of that particular class in the gold standard (labeled data set). In more specifically, Recall = TP / (TP + FN). Similarly, Accuracy of a class was also defined as the ratio between the numbers of correctly classified sequences and a total number of sequences for that particular class in the gold standard. However, F-measure can be computed as the harmonic mean of precision and recall, or 2 x Precision x Recall / (Precision + Recall).

Results

Experimental evaluation of the classification of utterance sequences by using the proposed probabilistic models is reported with two set of experiments: the first set of experiment was focused on natural sequences, while the second set of experiment was performed on alternating sequences.

Predictive performance for natural PPC code sequences

Predictive performance summary of the proposed methods on natural sequences is presented in Table 3. It follows that first-order HMM demonstrates the best performance achieving the highest precision 0.7980 with 0.7989 F-measure. However, second-order Markov Chain demonstrates the lowest performance in terms of precision and F-measure. It can be seen that the lowest recall of 0.7686 is achieved by the first-order MC. On the other hand, second-order HMM achieves the highest recall 0.8449 and accuracy 0.8449. From Table 3, it follows that second-order MC and HMM have lower precision and F-measure, but higher recall and accuracy than first-order MC and HMM. Precision and F-measure decrease by 6.74% and 2.39%, in case of MC, and by 7.27% & 2.09%, in case of HMM, when going from

first- to second-order models. In contrast, recall improves by 3%–5%, in case of both MC and HMM, when going from first- to second-order models. Overall, HMM performed better than MC for natural sequences.

Table 3: Performance of MC and HMM for predicting success of natural patient-provider communication sequences. The largest value for each performance metric is highlighted in bold.

Model	Order	Precision	Recall	F-Measure	Accuracy
Markov Chain	First order	0.7387	0.7686	0.7532	0.7686
	Second order	0.6889	0.7889	0.7352	0.7889
Hidden Markov Model	First order	**0.7980**	0.8059	**0.7989**	0.8059
	Second order	0.7400	**0.8449**	0.7822	**0.8449**

Predictive performance for alternating PPC code sequences

Table 4 summarizes predictive performance of the proposed methods on alternating sequences. These results indicate that second-order HMM had better recall than MC, achieving 97.13% recall. Similar to the case of natural sequences, first-order MC has the lowest performance compared to all other models across all metrics. Also, as in the case of natural sequences, prediction method based on second-order HMM has the highest recall of 0.9713 and the highest accuracy of 0.9713 among all models. However, different from natural sequences, second-order MC demonstrates the highest precision 0.9778 as well as the highest F-measure of 0.9736. We also observed that second-order MC and HMM outperform first-order MC and HMM. Specifically, moving from first-order to second-order MC improves precision, recall and F-measure of the MI success prediction method by 0.01%, 2.72% and 1.37%, respectively. Moving from first-order to second-order HMM improves precision, recall and F-measure of the MI success prediction method by 3.32%, 4.30% and 3.70%, respectively.

Table 4: Performance of MC and HMM for predicting success of natural patient-provider communication sequences. The largest value for each performance metric is highlighted in bold.

Model	Order	Precision	Recall	F-Measure	Accuracy
Markov Chain	First order	0.9777	0.9437	0.9604	0.9437
	Second order	**0.9778**	0.9694	**0.9736**	0.9694
Hidden Markov Model	First order	0.9392	0.9313	0.9353	0.9392
	Second order	0.9704	**0.9713**	0.9699	**0.9713**

Impact of hidden states

Since the number of hidden states may impact the classification performance, we use an optimum number of hidden states in HMMs. Figure 2 shows the performance of HMMs with a different number of hidden states. The results illustrate that, in general, the state number didn't impact the classification performance significantly. All four models show their best performance when state number in between 4-8. Table 5 shows an example of the distribution of PPC codes over the states for successful (S) and unsuccessful (U) HMMs. Log probability of each code reflects the state significance. In other words, these probabilities indicate that which PPC codes are close to adolescent commitment language and which are not?

Common patterns

Table 6 provides examples of typical PPC sequences that frequently appear in successful and unsuccessful motivational interviews. It can be seen that most successful patterns start with a summary of the discussion or open-ended/close-ended question. After that, if adolescents show positive behavior change, the counselor immediately reflects on that positive tendency, to reinforce adolescent's desire to adhere to positive behavior change. On the other hand, general information by the counselor can lead to a negative behavior change in adolescents, even in the cases when adolescents
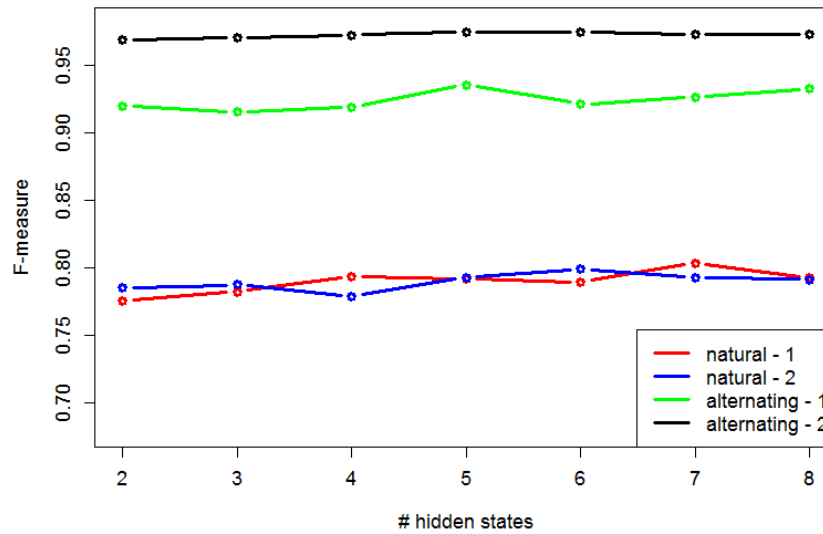
Figure 2: Performance of HMMs with different sequence type, order, and number of hidden states used in the experiment.

were showing positive tendency in their previous communication. This observation can be explained by adolescents quickly loosing focus when presented with general information.

Discussion

The purpose of this study was to facilitate the design of motivational interview by examining the performance of sequence classification in the field of health informatics. Studying the results across the different classification schemes we can make some key observations.

First, higher order models have better classification accuracy as compared to lower order models because they are able to capture longer ordering constraints present in the data set. However, the number of states in higher order Markov chains grows exponentially with the order of the chain. Consequently, it is harder to accurately estimate all the transition probabilities from the training set. That is, there are many high-order states that are not frequently observed in the training set, leading to inaccurate probability estimates. In principle, this problem can be solved if the size of the training set is increased at the same rate as the state-size of the higher order Markov chain. However, this cannot be easily done for many applications areas include our case. To skip these problems we restricted our study up to second order. The shorter average length of utterance sequence is another reason for keeping the order lower.

Second, the overall accuracy of hidden Markov model is better than general Markov model in both types of sequences. We observed that hidden Markov model achieved near-human accuracy to categories the sequences of behavior codes. This means that hidden states capture the information passes through the codes of the utterance sequence for determining the final outcome.

Third, alternating sequences provide higher performance for all Markov models. Even it provides the highest precision and F-measure for general Markov model. This indicates that each pair of alternating behavior codes forward their mutual information to the next step more accurately which improve performance for distinguishing outcomes of the motivational interview. And this is obvious because if we know the previous two behavior codes, for instance, provider's question and adolescent's response then it is easier to know the expected response for the current question of the counselor.

Table 5: Distribution of PPC codes over the states for successful (S) and unsuccessful (U) HMMs. Significant PPC codes are given in short form, which have log probability higher than -2.

State	Log probability of each symbol (301,LUP+, 327,OQ-TBN,101,EA,R-CT+,OQ-ECT+,AR,SUM-S, R-CT-S,RBA-S,G-INFO+,AF,314,CQ-EF,119,R-CML+, OQ-ECML+, 307, 325, 344,CHT-,106, OQ-ECT-, HUP-O, 345, 309, R-CT-S, 114, 304, 115, 329, 330, 103, 116, 324, 323, 118, 303, 334)
S-1	-2.896, -4.595, -4.031, -5.584, -11.383, -2.772, -1.532, -2.78, -3.816, -2.745, -8.444, -4.42, -3.485, -2.438, -8, -5.577, -10.333, -2.216, -3.462, -3.479, -3.71, -5.615, -20.984, -7.635, -5.7, -5.694, -3.629, -3.32, -3.195, -8, -5.007, -8, -4.636, -3.67, -8, -8, -11.027, -10.448, -8, -8, -8
S-2	-3.092, -1.344, -4.322, -4.707, -6.065, -2.205, -5.701, -2.064, -5.71, -3.778, -3.121, -6.97, -3.074, -3.505, -8, -3.879, -3.229, -5.707, -2.708, -4.729, -5.538, -4.993, -4.116, -3.125, -8, -3.29, -4.439, -4.414, -6.341, -8, -8, -8, -5.964, -7.009, -8, -8, -6.067, -6.071, -8, -8, -8
U-1	-8, -24.368, -8, -2.226, -8, -2.649, -1.458, -4.275, -2.632, -2.409, -54.292, -2.919, -2.867, -3.325, -8, -7.464, -31.322, -2.649, -3.022, -4.018, -8, -3.313, -51.78, -50.295, -2.919, -29.011, -8, -6.055, -3.341, -8, -8, -8, -24.716, -8, -8, -8, -8, -8, -8, -8
U-2	-8, -1.021, -8, -8, -8, -3.067, -6.497, -2.993, -8, -10.079, -3.794, -8, -2.448, -8, -8, -2.415, -3.1, -6.451, -5.018, -8, -8, -3.113, -2.407, -3.794, -8, -2.407, -8, -3.933, -7.201, -8, -8, -8, -3.794, -8, -8, -8, -8, -8, -8, -8, -8

Table 6: Common patterns in successful and unsuccessful motivational interviews.

Type	Pattern
successful	328 SUM-S: Summarize → 117 LUP+: Low Uptake, positive → 313 R-CML+: Reflect, commitment language positive → [positive commitment]
successful	307 SPT: Support → 117 LUP+: Low Uptake, positive → 313 R-CML+: Reflect, commitment language positive → [positive commitment]
successful	306 CQ-EF: Closed question, Elicit Feedback → 120 HUP-O: High Uptake, other → 313 R-CML+: Reflect, commitment language positive → [positive commitment]
unsuccessful	311 R-CT+: Reflect, change talk positive → 117 LUP+: Low Uptake, positive → 302 G-INFO+: General Information, positive → [negative commitment]
unsuccessful	302 G-INFO+: General Information, positive → 117 LUP+: Low Uptake, positive → 302 G-INFO+: General Information, positive → [negative commitment]
unsuccessful	305 EA: Emphasize Autonomy → 117 LUP+: Low Uptake, positive → 302 G-INFO+: General Information, positive → [negative commitment]

Finally, it was seen that successful interviews are more frequently responded by the adolescent with positive behavior changes which lead to making a positive commitment for loosing weight. Whereas, unsuccessful interviews are responded with negative behavior changes which lead to a negative commitment or no commitment.

Conclusion

In this paper, we proposed a method for based on Markov Chain and Hidden Markov Model could be used to build classifiers for the sequence of behavior codes in motivational interviews. Experimental results show that the behavior codes carry some inherited information to distinguish motivational interviews. This work can facilitate researchers to establish a causal relationship between different communication strategies and desired behavioral outcomes without having to repeatedly wade through pages of interview transcripts. It provides information that can directly inform and increase the efficiency of clinical practice for a successful interview. This work also explains some successful patterns for the practical use by the clinician during the interview session which will translate into positive change talk and commitment language. The direction of future work include the exploration of other methods such as BLAST[26] and Strand[27] for the classification of utterance sequences. We also have the plan to use large data set in conjunction with some advanced features for the evaluation of sequence classifiers.

Acknowledgments

References

[1] Leslie C, Kuang R. Fast string kernels using inexact matching for protein sequences. Journal of Machine Learning Research. 2004;5(Nov):1435–1455.

[2] Chuzhanova NA, Jones AJ, Margetts S. Feature selection for genetic sequence classification. Bioinformatics. 1998;14(2):139–143.

[3] Ye L, Keogh E. Time series shapelets: a new primitive for data mining. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 2009. p. 947–956.

[4] Kudenko D, Hirsh H. Feature generation for sequence categorization. In: AAAI/IAAI; 1998. p. 733–738.

[5] Lesh N, Zaki MJ, Ogihara M. Mining features for sequence classification. In: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 1999. p. 342–346.

[6] Nallam K, et al. An Effective Implementation of Hierarchical Approach For Sequence Classification. IJACTA. 2016;3(2):143–146.

[7] Keogh E, Kasetty S. On the need for time series data mining benchmarks: a survey and empirical demonstration. Data Mining and knowledge discovery. 2003;7(4):349–371.

[8] Keogh EJ, Pazzani MJ. Scaling up dynamic time warping for datamining applications. In: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 2000. p. 285–289.

[9] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE. 1989;77(2):257–286.

[10] Yakhnenko O, Silvescu A, Honavar V. Discriminatively trained markov model for sequence classification. In: Data Mining, Fifth IEEE International Conference on. IEEE; 2005. p. 8–pp.

[11] Deshpande M, Karypis G. Evaluation of techniques for classifying biological sequences. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer; 2002. p. 417–431.

[12] Wei L, Keogh E. Semi-supervised time series classification. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 2006. p. 748–753.

[13] Lane T, Brodley CE. Temporal sequence learning and data reduction for anomaly detection. ACM Transactions on Information and System Security (TISSEC). 1999;2(3):295–331.

[14] Drew J, Hahsler M, Moore T. Polymorphic malware detection using sequence classification methods and ensembles. EURASIP Journal on Information Security. 2017;2017(1):2.

[15] Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010. Jama. 2012;307(5):483–490.

[16] General US. Surgeon Generals vision for a healthy and fit nation. Washington, DC: HHS. 2010;.

[17] Kotov A, Hasan M, Carcone A, Dong M, Naar-King S, BroganHartlieb K. Interpretable probabilistic latent variable models for automatic annotation of clinical text. In: AMIA Annual Symposium Proceedings. vol. 2015. American Medical Informatics Association; 2015. p. 785.

[18] Hasan M, Kotov A, Carcone AI, Dong M, Naar S, Hartlieb KB. A study of the effectiveness of machine learning methods for classification of clinical interview fragments into a large number of categories. Journal of biomedical informatics. 2016;62:21–31.

[19] Carcone AI, Naar-King S, Brogan K, Albrecht T, Barton E, Foster T, et al. Provider communication behaviors that predict motivation to change in black adolescents with obesity. Journal of developmental and behavioral pediatrics: JDBP. 2013;34(8):599.

[20] Eide H, Quera V, Graugaard P, Finset A. Physician–patient dialogue surrounding patients expression of concern: applying sequence analysis to RIAS. Social Science & Medicine. 2004;59(1):145–155.

[21] Mutsam N, Pernkopf F. Maximum margin hidden Markov models for sequence classification. Pattern Recognition Letters. 2016;77:14–20.

[22] Eickeler S, Kosmala A, Rigoll G. Hidden markov model based continuous online gesture recognition. In: Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on. vol. 2. IEEE; 1998. p. 1206–1208.

[23] Srivastava PK, Desai DK, Nandi S, Lynn AM. HMM-ModE–Improved classification using profile hidden Markov models by optimising the discrimination threshold and modifying emission probabilities with negative training sequences. BMC bioinformatics. 2007;8(1):104.

[24] Won KJ, Prügel-Bennett A, Krogh A. Training HMM structure with genetic algorithm for biological sequence analysis. Bioinformatics. 2004;20(18):3613–3619.

[25] Chai W, Vercoe B. Folk music classification using hidden Markov models. In: Proceedings of International Conference on Artificial Intelligence. vol. 6. sn; 2001. .

[26] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990;215(3):403–410.

[27] Drew J, Hahsler M. Strand: fast sequence comparison using mapreduce and locality sensitive hashing. In: Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics. ACM; 2014. p. 506–513.