# Machine Learning Models for the Segmentation of eCoaching Text

**Mehedi Hasan, BS[1*], Alexander Kotov, PhD[1*], April Idalski Carcone, PhD[2], Ming Dong, PhD[1], Sylvie Naar, PhD[2]**
**[1]Department of Computer Science, Wayne State University, Detroit, Michigan**
**[2]Department of Family Medicine and Public Health Sciences, School of Medicine, Wayne State University, Detroit, Michigan**

**Abstract** *Poor eating habits, particularly low fruit and vegetable intake, is a growing, serious public health concern among young adults. An effective intervention is required to improve eating habits. eCoaching is an email-based intervention technique where a critical step is the segmentation of text for the automatic annotation of email exchange. In this study, we transformed this task into classification of detecting boundary of segmentation, and developed sevarals state-of-the-art machine learning models including Support Vector Machine, Naive Bayes, K-Nearest Neighbor (KNN), Recurrent Neural Networks by utlizing contextual, topic and punctuation mark features. Results indicate that KNN is the best model and achieved 0.986 F1-score in overall, 0.779 and 0.993 F1-scores for detecting boundary, not boundary, respectively. This study has a great implication to identify individual text segments, which can be annotated directly with a classification model, and accelerate the pace of identifying effective communication strategies linked to healthy eating.*

## Introduction

Unhealthy eating habits, particularly low fruit and vegetable intake, is a growing, serious public health concern, particularly among young adults age 21-30, referred to as Generation Y (GenY)[1,2]. This generation has adopted a lifestyle that involves eating accessible, "no mess", quick, "grab and go" foods[3,4]. They mainly eat "out" and infrequently shop and prepare food, limiting access to fruit and vegetables (FV)[5,6]. Unfortunately, less than one-third of US adults[1,7] and only 20% of GenY[1,8,9] eat the recommended 5 servings of fruit and vegetables daily. Those in inner city urban and rural settings have among the poorest eating habits[1,2,7–9]. GenY's poor dietary practices placing them at high risk for obesity and many chronic diseases, such as type 2 diabetes, as well as declines in predicted health status and life expectancy. Thus, there is a need to develop effective interventions to improve GenY's eating habits.

GenY is a tech-savvy generation requiring an intervention matched to their mobile lifestyle. Growing numbers use the internet to access health information with the largest increases in internet access among low-income Americans, making the internet well-suited for health promotion intervention[10]. MENU GenY[11] (Making Effective Nutrition Choices for Generation Y) is a technology-based public health intervention to encourage increased fruit and vegetable intake among GenY. A critical component of MENU GenY is personalized eCoaching. eCoaches use email to deliver motivation-enhancing coaching to encourage healthy eating, grounded in the principles of Motivational Interviewing (MI), an evidence-based communication technique to increase intrinsic motivation and self-efficacy for behavior change[12–14]. Patient "change talk", statements of intrinsic motivation about their desire, ability, reasons, need for and commitment to behavior change, is an established mediator of health behavior change[15]. Identifying specific communication strategies linked to behavior change and integrating these strategies into communication-based interventions (e.g., brief, motivation-enhancing interventions delivered in a variety of settings or public health initiatives) can increase these interventions' potency.

A major drawback of this research is the qualitative methods traditionally used to analyze the communication process which are resource-intensive, requiring an iterative process of human (subjective) interpretation of text. Rapidly developing computational technologies, specifically machine learning combined with classification models, offer a unique opportunity to accelerate this process. Our research group has recently applied machine learning-based models to similar communication data[16,17]. A simple communication code scheme was automated to characterize patient communication and achieved accuracy comparable to human coders[16]. The ultimate goal of the research study is to leverage innovative machine learning models to fully automate the communication coding process in eCoach-patient communication to increases in fruit and vegetable intake.

[*]Authors provided equal contribution.

However, a significant barrier of fully automate eCoaching is the unsegmented text data. Developing an automatic classification of clinical interactions required segmented text. Nevertheless, eCoaching data comprised of email responses which need to be segmented into group of MI behavior referes to "block of text". Automatic segementation of eCoaching intervention sessions is a challenging task due to the 2 important reasons. First, the email is an unstructured text that contains informal email exchange in non-traditional formats. Second, a text segment not necessarily belongs to the entire sentence or collection of sentences. One sentence can be segmented into several MI behaviors, and vice versa. Figure 1 illustrates the segmentation of an eCoaching email exchanage where first sentence segmented into 2 different MI behaviors. On the other hand, fourth and fifth segments contain only one and multiple sentences, respectively.

*On Mon Nov 10 20:40:02 2014, XXX wrote:*
*Hi YYY, I haven't had a chance to look through MD 5 or 6, but I've found a few veggies that I like to pack and take with me. I just have to prep them more. Thanks XXX*

(Email Date: 2014-11-11 10:29:18)
Hi XXX,

It's good to hear from you. It sounds like you found a plan that works for you as long as you are able to find time to prep veggies for on-the-go snacks. Sometimes people find inspiration for making a change by considering things that are important to them. There is some evidence that behavior change is often easier when it relates to your own values and goals. This might be helpful in finding reasons to keep up with what you are now doing. You stated that being considerate, respected, and responsible are important to you. How, if at all, would you say that eating better and having more energy would help you be considerate and respected? How about to be more responsible?

I look forward to hearing from you again soon,

YYY

**Figure 1: Segmentation of eCoaching text depicts the main challenges of boundary detection**.

In this paper, we address this problem by developing several state-of-the-art machine learning based models for the segmentation of eCoaching text to promote the automatic identification of best communication strategies without human interference. More specifically, we develop Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbor (KNN), Long Short Term Memory (LSTM), and Gated Recurrent Unit (GRU) by utlizing contextual, topic and punctuation mark features, to find the best model for the segmentation of eCoaching text.

Previous studies mainly focus on segemntation of text into sections and headers[18–21] or sentence boundary detection[22–24] in the medical domain. Apostolova et al.[18] applied SVM by utilizing word-vector cosine similarity metric combined with several heuristics to classify clinical report into semantic sections such as demographics, history, exam procedure, finding, impression, etc. After identification of each line in the document, Tepper et al.[20] trained an Maximum Entrophy models for the section classification. In 2009, Denny et al.[19] proposed a SecTag algorithm, which combined natural language processing technique, terminology-based rule, and naive bayesian score for identifying sections and headers that achieved 99% recall with 95.6% precision. On the other hand, SVM exploiting with linear kernel and recurrent convolutional neural networks with posodic, part of speech features and word embeddings, were trained by Kreuzthaler et al.[23] and Griffis et al.[22], respectively, for the detection of sentence boundary. However, segmentation of clinical text, in partiicular, segmentation of MI or eCoaching text into group of MI behavior is ignored while relying on manual hand-coded approach. Therefore, this study introduce a novel approach and the authors are not aware of any other work this approach has been considered for the segmentation of MI or eCoaching text into "block of text".

## Methods

### *Data collection*

The experimental dataset for this work was constructed from the 49 eCoaching sessions, which include a total of 3,138 segmented and annotated MI behaviors. Each session contains an MI intervention involving patient-provider communications in email. To filter out noise from the dataset, non-ascii characters are removed and then applied stemming to obtain a general form of word from different word representations, such as “eating”, “eats”, and “eat”. We formulate the text segmentation task into a binary classification, as shown in Figure 2. An intervention session with email exchange is given as the input, it is partitioned into adjacent word pairs by sliding them. Each pair of them classified into either of the two categories: “boundary” and “not boundary”. The text is segmented at position, where an adjacent word pair classified into “boundary” class. If all pairs of word classified into “not boundary”, the text is treated as one about a single MI behavior. Totaly, we obtained 95,421 word pairs, which include 3,138 “boundary” and 92,283 “not boundary” instances.

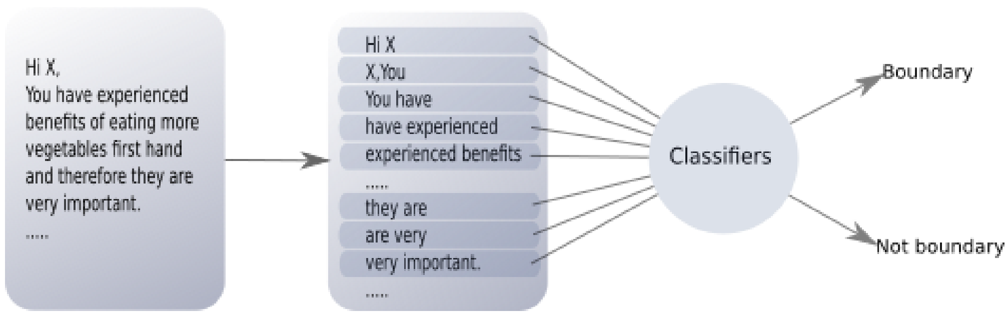


**Figure 2: Transformation of text segmentation task into text classification task**.

For the experiment, we utilized three type of features including word (textual feature), tpoic, and punctuation mark. Each word represented in a binary format, where 1 indicates the appearance of the word and 0 for absence. Topics are considered as features since topic models are very effective[17,25,26] to represent text documents. In this paper, we exploit the Labeled LDA model[17] and represent each word in a vector of 2 topics, where the number of topics is experimentaly determined by the model performance. Punctuation mark containing one of the symbols {‘.’, ‘,’, ‘!’, ‘?’, ‘:’, ‘;’, ‘-’} is also employed as feature. This is one of the most important feature as they indicate the boundary of a sentence, clause and phrase.

### *Segmentation classifiers*

Several state-of-the-art classifiers, including Naive Bayes (NB)[27], Support Vector Machine (SVM)[28], K-Nearest Neighbour (KNN)[27], two variant of Recurrent Neural Networks (RNN)[29]: Long Short Term Memory (LSTM)[30] and Gated Recurrent Unit (GRU)[31], are employed to estimate the classification performance.

**Naive Bayes**: this model is constructed by using the training data and estimate the prior probability of classes, and each feature given the class. Then, posterior probability is computed to predict the class label by applying the bayes theorem with the assumption that features are conditionally independent. This study utilized a specialized version of Naive Bayes called Multinomial Naive Bayes, which is best suitable for discrete features such as word.

**Support Vector Machine**: we used this model as one of the state-of-the-art classification technique proven to perform well in text categorisation[32] for its ability to cope with very high dmensional input feature space. SVM finds the best hyperplane in the feature space that maximize the separation between the closest “boundary” and “not boundary” training examples. In this experiment, polynomial kernel is employed to train the SVM model for the segmentation of eCoaching text.

**K-Nearest Neighbour**: By this model, each training sample represented as a point in the input feature space. For a new test sample, Euclidean distance is calculated to find the k-nearest neighbors. Finally, the test sample is classified into majority class of the k-nearest neighbors. We experimentaly determined that best performance was achieved with k = 3 for the classification of word pairs.

**Recurrent Neural Networks**: RNN is a neural network architecture designed to capture sequential patterns present in temporal sequence such as text data. When we predict the "boundary" point, adjacent word pair will help to understand the pattern of the sequence. Long Short Term Memory networks usually reffered as LSTMs[30], are a special type of RNN capable of handling variable size input sequence, contains internal memory. GRU[31] is a variant of LSTM mathematically represented by the following formula:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{1}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{2}$$

$$\tilde{h}_t = tanh(W_h x_t + r_t \odot U_h h_{t-1} + b_h) \tag{3}$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \tag{4}$$

In Eq. 1-4, $\sigma$ corresponds to sigmoid function and $\odot$ designates an element-wise product. The update gate $z_t$ and reset gate $r_t$ at time step $t$ are computed by the Eq. (1) and (2), where $W_z$, $W_r$, $W_h$, $U_z$, $U_r$, $U_h$ are the weight matrices and $b_z$, $b_h$ and $b_r$ are bias vectors. The activation $h_t$ of the GRU at time $t$ is a linear combination of previous activation $h_{t-1}$ and the candidate activation $\tilde{h}_t$, which is represented by Eq. (4) and (3). We build our RNN model with one hidden layer, output layer, and input layer which get one hot encoding of word vector as input. Since one-hot vector is given in the input layer, results are reported with textual features and punctuation marks only. We experimentally determined that the best performance is achieved when the number of hidden units = 32, batch size = 8, optimizer = adam, as well as 600 epochs is used based on the validation loss.

### *Evaluation metrics*

In this experiment, standard metrics: precision, recall, and F-measure, are applied to evaluate the performance of binary classifiers[33]. However, we didn't report accuracy as a performance metric because accuracy is highly sensitive to the prior class probabilities and does not fully describe the actual difficulty of the decision problem for unbalanced dataset. We conduct the experiment with 5 folds cross-validation and weighted macro-averaging of these metrics over the folds. All models are trained on 80% of the word pairs and remaining 20% of the data is used as a test set for reporting the performance of the model. We also estimated the area under the receiving operating characterstics (ROC) curve[34] (AUC) metric due to its effectiveness in measuring the quality of binary classifiers for imbalanced datasets[35].

## Results

opening para: evaluated with boundary, not boudary and overall

**Table 1:** Performance of NB, SVM, KNN, and RNN for detecting boundaries of segmentation in eCoaching text. The highest value for each performance metric is highlighted in bold.

| **Method** | **contextual features only** | | | **contextual + punctuation marks (+ topics except RNN)** | | |
|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1-Score** | **Precision** | **Recall** | **F1-Score** |
| NB | 0.594 | 0.662 | 0.626 | 0.590 | 0.666 | 0.626 |
| SVM | 0.742 | **0.679** | 0.709 | 0.774 | 0.696 | 0.733 |
| KNN | **0.808** | 0.663 | **0.728** | **0.820** | **0.742** | **0.779** |
| LSTM | – | – | – | 0.619 | 0.416 | 0.497 |
| GRU | – | – | – | 0.642 | 0.490 | 0.554 |

para 1: 0) opening with referencing table 1) KNN best in all metrics except Recall 2) RNN shows lowest performance in terms of recall and F1 score, although GRU better than LSTM 3) SVM is second highest model and highest model in terms of recall when only textual features is used. 4) NB shows lowest precision value among all and better than

RNN with recall and F1-score 5) for additional features, recall and F1 increase x, y, z, for a, b, c, respectively 6) However, precision increase for all except NB

**Table 2:** Performance of NB, SVM, KNN, and RNN for detecting no boundaries in eCoaching text. The highest value for each performance metric is highlighted in bold.

| **Method** | **contextual features only** | | | **contextual + punctuation marks (+ topics except RNN)** | | |
|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1-Score** | **Precision** | **Recall** | **F1-Score** |
| NB | 0.988 | 0.985 | 0.987 | 0.989 | 0.984 | 0.986 |
| SVM | **0.989** | 0.992 | 0.991 | 0.990 | 0.993 | 0.991 |
| KNN | **0.989** | **0.995** | **0.992** | **0.991** | **0.994** | **0.993** |
| LSTM | – | – | – | 0.981 | 0.991 | 0.986 |
| GRU | – | – | – | 0.983 | 0.991 | 0.987 |

para 2: 0) opening with referencing table 2) Similar to boundary detection, KNN is best and RNN is lowest in all metrics for 'not boundary' identification 3) Performance is very high compared to boundary detection however, perform better than random 4) Effect of additional feature is consistent in 'not boundary' classification. 5) for additional features, precision, recall and F1 increase x, y, z, for a, b, c, respectively

**Table 3:** Weighted average performance of NB, SVM, KNN, and RNN for the segmentation of eCoaching text in detecting both "boundary" and "not boundary". The highest value for each performance metric is highlighted in bold.

| **Method** | **contextual features only** | | | **contextual + punctuation marks (+ topics except RNN)** | | |
|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1-Score** | **Precision** | **Recall** | **F1-Score** |
| NB | 0.975 | 0.974 | 0.975 | 0.976 | 0.974 | 0.975 |
| SVM | 0.981 | 0.982 | 0.981 | 0.983 | 0.983 | 0.983 |
| KNN | **0.983** | **0.984** | **0.983** | **0.986** | **0.986** | **0.986** |
| LSTM | – | – | – | 0.969 | 0.972 | 0.970 |
| GRU | – | – | – | 0.972 | 0.974 | 0.973 |

para 3: 0) opening with referencing table 1) Overall, KNN is best and RNN is lowest in all metrics for 'not boundary' identification 2) SVM and RNN achieved x and xx, y and yy, z and zz for precision, recall and F1, respectively. 3) Effect of additional feature is also consistent. 4) for additional features, precision, recall and F1 increase x, y, z, for a, b, c, respectively

## Discussion

This study is the first large-scale efforts to evaluate the segmentation of eCoaching text. Experimental results indicate that KNN is the best model among all machine learning methods considered for this study. KNN achieved 0.986 F1-score in overall, 0.779 and 0.993 F1-scores for detecting "boundary" and "not boundary", respectively. Robust performance of KNN provides the evidence that machine learning models are capable to learn information from the email exchange. Although the domain of this study was intentionally quite small, we believe that our study is not limited to ecoacing domain, and it can be successfully applied to other domain as well.

The additional topic and punctuation mark features, made significant improvement in performance of all maching learning methods. In every cases, model performs better when contextual feature is used in combination with topic and punctuation mark features. Thsi results also mean that segmentation performance might be improved by adding more relevant features including human insight into the problem.

In this paper, results are reported by each class to avoid confusion about the overall model performance. In addition, standard metrics: precision, recall and F1-score were used to eliminate doubt about the model performance because accuracy is misleading for imbalance dataset. AUC values are also outlined due to its effectiveness in measuring

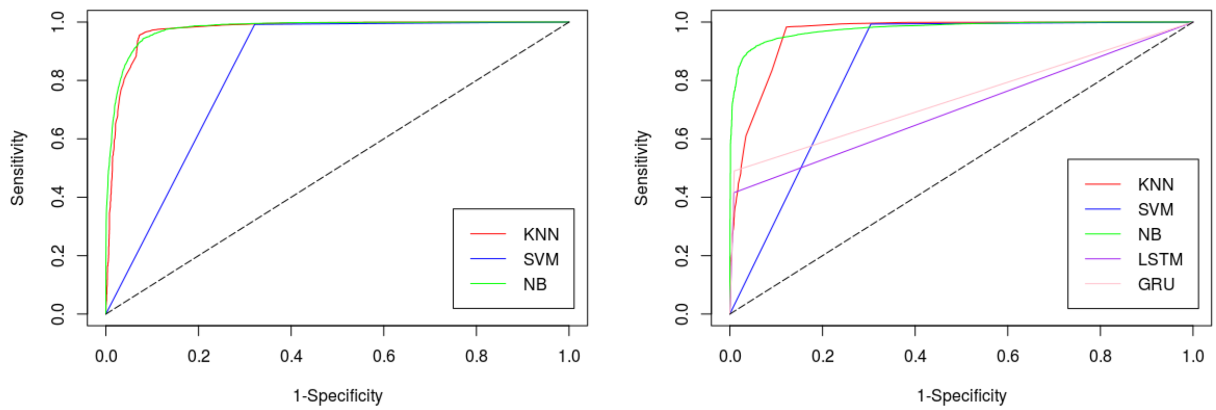


**Figure 3: Receiver operating characteristic curves showing the performance of binary classifiers for the segmentation of eCoaching text when textual features (left) and combination of textual and other features (right) are used**.

the quality of binary classifiers for imbalanced datasets[35], which was demonstrated by the ROC curves in Figure 3. NB shows the highest AUC values and achieved 0.978 for both cases while it provides lowest classification results except RNN. On the other hand, KNN and SVM exibits 0.972 and 0.835 AUCs when only textual features are used; and 0.959 and 0.844 AUCs when combination of textual, topic and punctuation mark features are used. Finally, RNN demonstrate lowest AUC values among all machine learning models, achieved AUC values 0.704 and 0.740 for LSTM and GRU, respectively.

We observed worst results of RNN, in particular LSTM and GRU for the text segmetation. We believe that RNN performed poorly because it has a large set of weights which required large set of data for both class. It this study, we utilized 3,138 examples of boundary case which failed to achieve good results. While performance of RNN is poor, GRU performed better than LSTM which was already observed in other previous study[36].

Punctuation mark plays an important role in segmentation boundary detection, and a large numbers of errors were encountered by the false positive of boundary identification. Similarly, additional information is the common reason of classified original segment into multiple segments. For example, [need help from April].

Our proposed approach is a novel approach for text segmentation where previous study focus on the segmentation of text into sections, headers, and sentences. However, this study segmented email exchange into “block of text” represent a particular MI behavior. This work will significantly reduce the amount of resource and time required to segemnt email text manually. Furthermore, this paper can help to annotate each segment automatically by building a new classifer, which will accelerate the pace of finding best communication strategies to develop an effective MI intervention for healthy eating.

As for future work, we plan to evaluate our methods with several similar and different datasets. We also plan to use combination of machine laerning and natural language approaches to improve model performance. For example, part-of-speech tagging and distance from boundary of the sentence might significantly enhance the classification performance. The limitation of this study is that eCoaching text are collected from a single medical institute; formatting, style, and email segment can be different in other settings.

## Conclusion

Segmentation of eCoaching text is an integral part of developing an automated eCoaching intervention. Although several studies have done for the segmentation of clinical text into sections and sentences, none of them are used for the segmentation of text into a group of MI behavior in the setting of discourse analysis with email under the principle of motivational interviews. In this paper, we compared the performance of machine learning models for the task of segmentation of e-coaching text. We found out that k-nearest neighbour provides the best performance for the segmentaion of text in terms of all performance metrics. Manual segmentation of e-coaching data is very resource-intensive and time consuming task, which can significantly decrease the time and effort required to develop effective behavioral interventions. Our proposed methods can help to identify individual text segments, which can be annotated directly with a classification model and increase the effectiveness of behavioral interventions. This approach will help for developing fully automated eCoaching and also accelerate the pace of identifying effective communication strategies linked to healthy eating. As our future work, we plan to evaluate our approach on other datasets involves in discourse analysis for enhancing our proposed method.

**Acknowledgments**

This study was supported by a grant from the National Institutes of Health, NIDDK R21DK108071, Carcone and Kotov, MPIs. We would like to thank the student assistants in the Department of Family Medicine and Public Health Sciences at Wayne State University School of Medicine for their help in developing the training dataset by manually segmented the eCoaching text.

**References**

[1] Blanck HM, Gillespie C, Kimmons JE, Seymour JD, Serdula MK. Trends in fruit and vegetable consumption among US men and women, 1994–2005. Preventing chronic disease. 2008;5(2).

[2] for Disease Control C, (CDC P, et al. Fruit and vegetable consumption among adults–United States, 2005. MMWR Morbidity and mortality weekly report. 2007;56(10):213.

[3] Nebeling L, Yaroch AL, Seymour JD, Kimmons J. Still not enough: can we achieve our goals for Americans to eat more fruits and vegetables in the future? American journal of preventive medicine. 2007;32(4):354–355.

[4] Brug J, Campbell M, van Assema P. The application and impact of computer-generated personalized nutrition education: a review of the literature. Patient education and counseling. 1999;36(2):145–156.

[5] Nelson MC, Lytle LA, Pasch KE. Improving literacy about energy-related issues: the need for a better understanding of the concepts behind energy intake and expenditure among adolescents and their parents. Journal of the American Dietetic Association. 2009;109(2):281–287.

[6] Larson NI, Perry CL, Story M, Neumark-Sztainer D. Food preparation by young adults is associated with better diet quality. Journal of the American dietetic association. 2006;106(12):2001–2007.

[7] Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. Jama. 2006;295(13):1549–1555.

[8] Association ACH, et al. American college health association national college health assessment (ACHA-NCHA) spring 2005 reference group data report (abridged). Journal of American College Health. 2006;55(1):5.

[9] Thompson TG, Veneman AM. Dietary guidelines for Americans 2005. United States Department of Health and Human Services and United States Department of Agriculture. 2005;.

[10] Strecher V. Internet methods for delivering behavioral and health-related interventions (eHealth). Annu Rev Clin Psychol. 2007;3:53–76.

[11] Alexander GL, Lindberg N, Firemark AL, Rukstalis MR, McMullen C. Motivations of Young Adults for Improving Dietary Choices: Focus Group Findings Prior to the MENU GenY Dietary Change Trial. Health Education & Behavior. 2017;p. 1090198117736347.

[12] Miller WR, Rollnick S. Motivational interviewing: Helping people change. Guilford press; 2012.

[13] Miller WR, Rollnick S. Ten things that motivational interviewing is not. Behavioural and cognitive psychotherapy. 2009;37(2):129–140.

[14] Miller WR, Rose GS. Toward a theory of motivational interviewing. American psychologist. 2009;64(6):527.

[15] Apodaca TR, Longabaugh R. Mechanisms of change in motivational interviewing: a review and preliminary evaluation of the evidence. Addiction. 2009;104(5):705–715.

[16] 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.

[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] Apostolova E, Channin DS, Demner-Fushman D, Furst J, Lytinen S, Raicu D. Automatic segmentation of clinical texts. In: Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE. IEEE; 2009. p. 5905–5908.

[19] Denny JC, Spickard III A, Johnson KB, Peterson NB, Peterson JF, Miller RA. Evaluation of a method to identify and categorize section headers in clinical documents. Journal of the American Medical Informatics Association. 2009;16(6):806–815.

[20] Tepper M, Capurro D, Xia F, Vanderwende L, Yetisgen-Yildiz M. Statistical Section Segmentation in Free-Text Clinical Records. In: LREC; 2012. p. 2001–2008.

[21] Cho PS, Taira RK, Kangarloo H. Text boundary detection of medical reports. In: Proceedings of the AMIA Symposium. American Medical Informatics Association; 2002. p. 998.

[22] Griffis D, Shivade C, Fosler-Lussier E, Lai AM. A quantitative and qualitative evaluation of sentence boundary detection for the clinical domain. AMIA Summits on Translational Science Proceedings. 2016;2016:88.

[23] Kreuzthaler M, Schulz S. Detection of sentence boundaries and abbreviations in clinical narratives. In: BMC medical informatics and decision making. vol. 15. BioMed Central; 2015. p. S4.

[24] Treviso MV, Shulby C, Aluísio SM. Sentence segmentation in narrative transcripts from neuropsychological tests using recurrent convolutional neural networks. arXiv preprint arXiv:161000211. 2016;.

[25] Hashimoto K, Kontonatsios G, Miwa M, Ananiadou S. Topic detection using paragraph vectors to support active learning in systematic reviews. Journal of biomedical informatics. 2016;62:59–65.

[26] Lu HM, Wei CP, Hsiao FY. Modeling healthcare data using multiple-channel latent Dirichlet allocation. Journal of biomedical informatics. 2016;60:210–223.

[27] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. Journal of machine learning research. 2011;12(Oct):2825–2830.

[28] Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM transactions on intelligent systems and technology (TIST). 2011;2(3):27.

[29] Bengio Y, Frasconi P, Simard P. The problem of learning long-term dependencies in recurrent networks. In: Neural Networks, 1993., IEEE International Conference on. IEEE; 1993. p. 1183–1188.

[30] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997;9(8):1735–1780.

[31] Cho K, Van Merriënboer B, Bahdanau D, Bengio Y. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:14091259. 2014;.

[32] Joachims T. Text categorization with support vector machines: Learning with many relevant features. In: European conference on machine learning. Springer; 1998. p. 137–142.

[33] Aas K, Eikvil L. Text categorisation: A survey. Technical report, Norwegian computing center; 1999.

[34] Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian pediatrics. 2011;48(4):277–287.

[35] Hu J, Yang H, King I, Lyu MR, So AMC. Kernelized Online Imbalanced Learning with Fixed Budgets. In: AAAI; 2015. p. 2666–2672.

[36] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:14123555. 2014;.