Emotion Classification and Textual Clustering Techniques for Gang Intervention Data

Ruofei Wu, Chenxin Yang, David Hyde, Andrea L. Bertozzi
Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90095
lilianwu@ucla.edu, chenxinyang@g.ucla.edu,
dabh@math.ucla.edu, bertozzi@math.ucla.edu

P. Jeffrey Brantingham
Department of Anthropology
University of California, Los Angeles
Los Angeles, CA 90095
branting@ucla.edu

***Abstract*—We study a recent dataset documenting the nature of gang involvement among 14-25 year-olds participating in the Los Angeles Mayor's Office of Gang Reduction and Youth Development (GRYD) Intervention Family Case Management Program. We use natural language processing techniques, including emotion classification and textual clustering, to perform quantitative analyses of free-form responses in the data. These analyses yield insights into the effectiveness of the program and provide a better understanding of its participants. We also compare several computational techniques and remark on their relative effectiveness in application to this dataset.**

***Index Terms*—gang membership, gang intervention, natural language processing, emotion classification, sentiment analysis**

I. Introduction

Gangs present longstanding challenges for many communities, particularly large urban centers [1]–[3]. Gang involvement and activity can pose threats to personal development, community health, and public safety [4]–[8]. Therefore, numerous cities and non-profit groups have developed programs for curtailing gang involvement and activity within their communities [9]–[13].

Since 2008, the City of Los Angeles Mayor's Office Gang Reduction & Youth Development Office has funded prevention and intervention services throughout the city as part of a comprehensive strategy to address gang violence [14]. The GRYD Intervention Family Case Management (FCM) Program is designed to increase youth and family protective factors and resiliency, while reducing gang embeddedness for gang-involved youth and young adults between ages 14–25. Participants who engage with GRYD FCM services are asked to complete the Social Embeddedness Tool (SET) questionnaire [15], which asks about family background, substance use, lifestyle, personality, gang involvement, gang-related activity, and many other attitudinal and behavioral features. The intent of the SET instrument is to evaluate how close an individual is to the center of the gang (i.e., gang embeddedness) and to record a broad collection of covariates. While many participants only complete the SET questionnaire once, others complete it at multiple points in time (typically every six-months in the program), giving insight as to how GRYD FCM services affects participants over time. While many of the questions in the SET rely on Likert-like scale responses [16], several questions yield open-ended, free-form, textual responses. Additionally, GRYD FCM Providers record the problems and strategies identified by participants and the FCM Strategy Team when building case plans. Both the SET data and data on participant problems and strategies to address the problem were used in this analysis.

In this paper, we apply natural language processing (NLP) techniques, such as emotion classification and textual clustering, to the SET free-text responses to produce novel insights into trends of gang involvement and activity. Our results suggest significant correlations between various factors such as detected emotions, the problems and strategies identified by GRYD FCM providers, and subsequent participant outcomes. As part of our work, we compare several state-of-the-art approaches for emotion classification, which continues to be an active area of NLP research.

II. Related Work

A. Gang Embededdness and Delinquency

Researchers often consider the notion of *gang embeddedness* when describing how an individual is affiliated with a gang. Gang embeddedness generally refers to how close someone is to the center of the gang. [17] uses a simple three-tiered classification for prison gang embeddedness. [18] uses personality measures and constructs related to personal resilience and antisocial personality in order to define a measure of gang embeddedness. Based on Hagan's notion of criminal embeddedness [19], [20] proposed a graded gang embeddedness model based on item response theory. Thus, embeddedness may be measured with various metrics, and an appropriate quantitative description of gang embeddedness largely depends on the data one has available.

Some studies have explored the consequences of gang embeddedness in terms of individual or group outcomes. For example, [21] observed that arresting a gang leader (someone with high embeddedness) can potentially result in temporarily increased gang violence. [22] studied how decreasing gang embeddedness (disengaging form gangs) affected personal

978-1-7281-6251-5/20/$31.00 ©2020 IEEE

social and developmental outcomes. Works like these inspire a related consideration of individual *risk scores*, capturing how likely it is that an individual will engage in delinquent behavior. We hypothesize that there is a correlation between gang embeddedness and (gang-affiliated) risk scores. Data collected by GRYD FCM providers allows us to directly examine such risk scores, and we explore what factors make an individual more or less at risk according to this metric.

B. Natural Language Processing

The field of NLP has made great advances in recent years in understanding certain aspects of free-form textual content. These advances were aided by both "the data explosion" [23] and the advent of deep learning. A particularly interesting area in NLP is sentiment analysis [24]–[26], which generally aims to understand what emotions are conveyed by text and with what intensity. A common example is understanding whether messages on Twitter convey positive or negative emotion, e.g. [27], [28].

A related but less common NLP task is emotion classification. Instead of scoring whether the sentiment of text is positive or negative, emotion classification seeks to label or score a text by the semantic emotions it conveys (joy, fear, anger, etc.). This is an inherently more challenging task since there are more possible labels and hence greater room for confusion/error; though recently, a growing number of researchers are addressing this challenge. [29] considered long short-term memory (LSTM) networks and BiLSTM approaches with attention mechanisms to classify emotions on a tweet data set. [30] proposed a dual attention-based transfer learning approach which uses sentiment analysis to improve results of emotion classification. [31] presented a method based on graph convolutional networks. In general, the maximum per-class accuracy in recent work hovers around 80%, meaning there is still ample room for novel algorithms.

III. Computing Risk Scores

To numerically evaluate a GRYD FCM participant's tendency to take risks and engage in delinquent behaviors, we calculate a *risk score* based on their answers to 20 questions in the SET questionnaire, related to gang embeddedness, delinquent behavior, and personality traits. Specifically, we group questions into categories according to four sociological factors: personal behavior norms (4 questions), impulsive risk taking (4 questions), experience with arrest and detention (10 questions), and experience with delinquency and violence (2 questions).

A person's score in each factor is calculated by taking the average of their answers to all related questions. Given these per-category scores for an individual, we then convert each of the four raw scores to a corresponding z-score by the formula `(score - mean)/standard deviation`. Summing the four z-scores yields our risk score, a composite score of the four equally-weighted factors. This metric is designed so that the lower the score, the less the person is at risk for delinquent (in particular, gang-related) activity. Our risk score allows for weighting of the different categories if, for example, experience with arrest is known to have greater correlation with future delinquency than impulsive risk taking; we leave tuning these component weights to future work.

IV. Emotion Classification

We are interested in characterizing emotional factors of GRYD FCM participants. We therefore perform emotion classification on an open-ended question that asks participants to briefly talk about stressful or upsetting things that have happened to them. As is common in the recent literature, we choose to pursue learning-based approaches rather than relying on limited ontological or linguistic models [32].

A. Data Selection

For training our models, we use a unified dataset [33] that combines news headlines, tweets, tales, conversation data, and other large-scale corpora [34]–[42]. Our dataset consists of 190,895 texts in total with a small set of standard emotion labels commonly used psychological studies, i.e., joy, neutral, anger, sadness and fear. Although answers from our SET question of interest mainly consist of negative events and positive emotion labels are expected to be rare, we keep the "joy" emotion in our model for the purpose of generalizing our trained models to other questions.

B. Traditional Machine Learning Methods

We randomly split the data into 80% training 20% testing data. We then evaluate traditional machine learning methods to classify texts using only one of the five emotional labels. We first clean the text data using tokenization, punctuation and stopword removal, stemming, and lemmatization. We then vectorize texts using `TF-IDF` from the `Scikit-Learn` library [43]. Words between given minimum and maximum document frequency scores constitute the features for training and test vectors; default thresholds and parameters in [43] were used except for specifying the usage of unigrams and bigrams, as well as specifying the usage of sublinear term frequency.

A range of ML methods including naïve Bayes, random forest, logistic regression, and support vector machines produced unsatisfactory results. (For Naïve Bayes, we consider two implementations: one from the `Scikit-Learn` library [43] and one from NLTK [44]. The latter does not use `TF-IDF` when preprocessing the data.) Although the all-class accuracies shown in Table I seem promising, a confusion matrix analysis reveals that prediction for individual emotions is not accurate. In fact, the lowest prediction accuracy on the test set for a particular emotion is 5%. Hence we focus on deep learning approaches for the remainder of this section.

C. Convolutional Neural Network Modeling

Convolutional neural network (CNN) models were originally developed for image classification and feature learning, in which the model accepts a two-dimensional input representing an image's pixels and color channels. This similar process can be applied to one-dimensional (1D) sequences of text data.

TABLE I
MULTI-CLASS ACCURACY FOR ALL LEARNING METHODS CONSIDERED.

Method	Accuracy
Naïve Bayes (Scikit-Learn)	55.95%
Naïve Bayes (NLTK)	68.2%
Random Forest	62.97%
Logistic Regression	66.68%
Support Vector Machine	67.86%
CNN	69.71%
GRU	70.25%
BERT	76.11%

In fact, a simple CNN with only one convolution layer can achieve great performance in deep learning for NLP [45]. Accordingly, we consider training a 1D CNN model in the present work.

The first step of building the CNN model is to generate vectors for each word. Many researches have shown that initializing word vectors with those obtained from an unsupervised neural language model can improve performance in the absence of a large supervised training set [46]. Thus, after preprocessing the data, we first create a word embedding layer on top of 300-dimensional `Word2Vec` word vectors that were pre-trained on Wikipedia articles [47]. Then, one layer of convolution is applied with 512 different filters to produce features. These filters are expected to be able to discern different emotion markers. For instance, one filter may specialize in detecting "joy," e.g. it may be strongly activated by textual signals like "great" or "better." The next operation is to apply a max pooling layer over the feature map to obtain the most important features. These important features are passed to a fully connected softmax layer to produce five emotion classes: neutral, joy, anger, fear and sadness. The general network structure is visualized in Figure 1; more details of CNN modeling are found in [45].

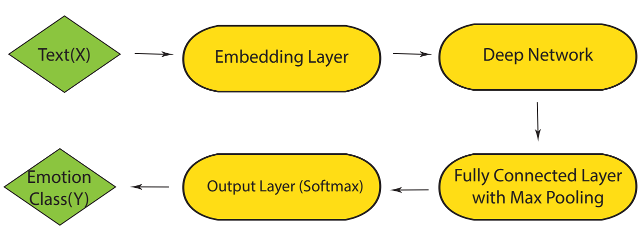

Fig. 1. Flow chart for CNN model.

With this CNN model, we achieve 69.71% overall classification accuracy on our validation set. The model performs particularly well on predicting the sadness class. The confusion matrix for all of the classes is shown in Figure 2.

D. Bidirectional GRU

The concept of a BiRNN (Bidirectional Recurrent Neural Network) was introduced to overcome limitations of a regular RNN [48]. The advantage of a BiRNN is that it can be trained using all available input information from both backwards and forward states. It has a structure that splits neurons of a regular RNN into two parts, which are responsible for positive time direction and negative time direction respectively [48]. The GRU (Gated Recurrent Unit), which was first used in the context of machine translation, is a useful recurrent unit. The most important feature of this approach is that it encodes the input sentence into a sequence of vectors and adaptively captures a subset of these vectors while decoding [49]. Long short-term memory (LSTM) [50] is another architecture for recurrent neural networks (RNNs), which have feedback (recurrent) connections as opposed to purely feedforward networks. LSTMs are a particularly effective type of RNN because they can overcome the vanishing gradient problem [51]. Both LSTM units and GRUs have been demonstrated to be superior to traditional recurrent units as the convergence is often faster, and the solutions tend to be better [52]. In our case, we trained a Bidirectional RNN model with a GRU layer[1].

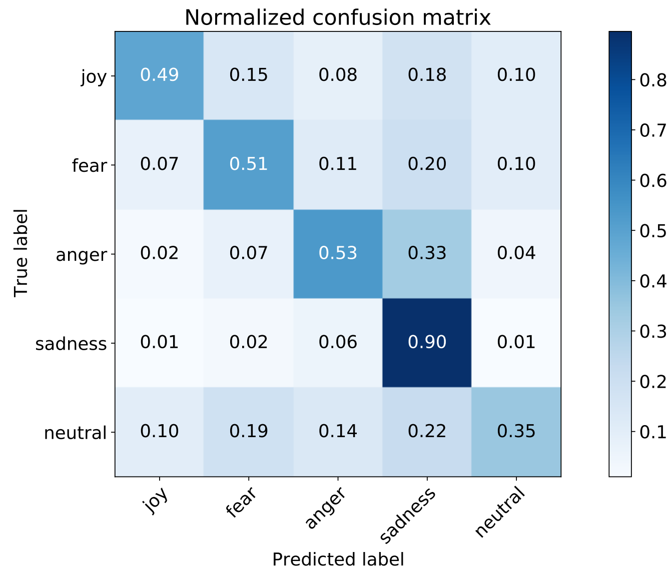

Fig. 2. Confusion matrix for CNN model.

The model built with a bidirectional GRU also achieves great performance levels on our training data with overall classification accuracy of 70.25%. In particular, it classifies sadness emotion quite well. For the validation set, 90% of sadness events are correctly classified.

E. BERT

BERT (Bidirectional Encoder Representations from Transformers) is a language representation model proposed in [53]. It is a state-of-the-art pre-trained model that can be used for a wide variety of tasks. By using a "masked language model" (MLM) pre-training objective which predicts the masked tokens from the input based on both the left and the right context, BERT enables pre-trained deep bidirectional representations. As a result, it achieves excellent performance on a broad range of token-level tasks including question answering. Therefore,

[1]https://github.com/lukasgarbas/nlp-text-emotion

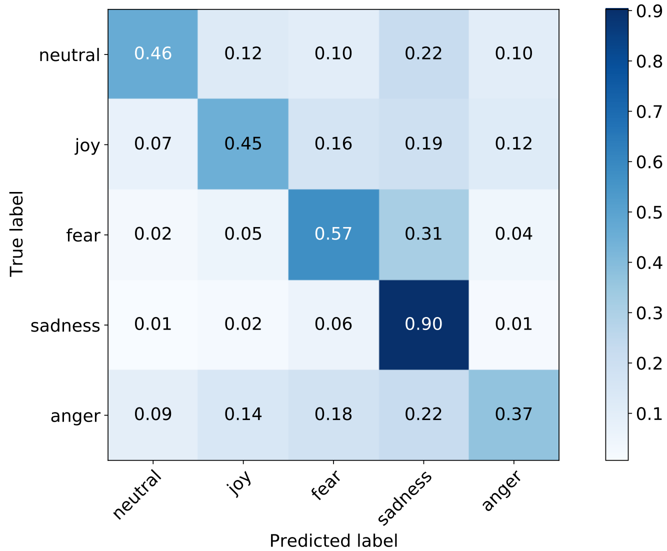

Fig. 3. Confusion matrix for Bidirectional GRU model.

we also apply BERT to our problem of multiclass emotion classification. The classification process includes preprocessing of textual data, training, and validation with the pre-trained BERT model[2].

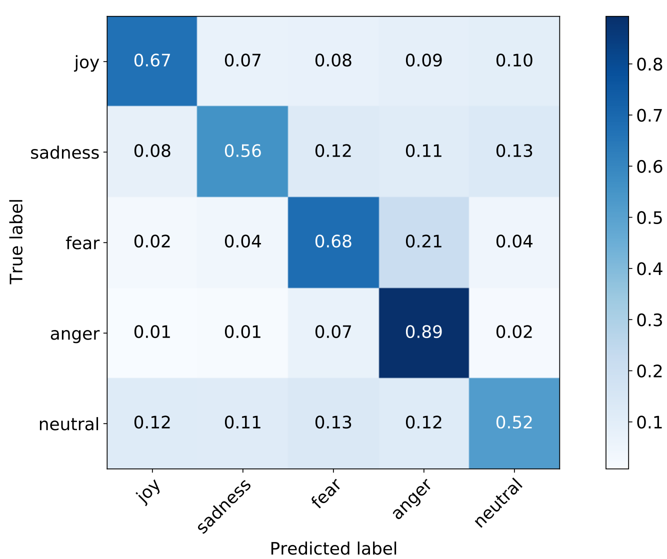

Fig. 4. Confusion matrix for BERT model.

The overall classification accuracy achieved by training with a BERT model is 76.11%. Figure 4 shows the confusion matrix for the results obtained using BERT. As we can observe from the confusion matrix, BERT has the best performance in terms of classification accuracy for each emotion. All of the emotions are classified with an accuracy over 50%, and anger has the highest classification accuracy with up to 89%.

[2]The model used is https://tfhub.dev/google/bert_uncased_L-12_H-768_A-12/1, trained on Wikipedia and the BookCorpus, see [53].

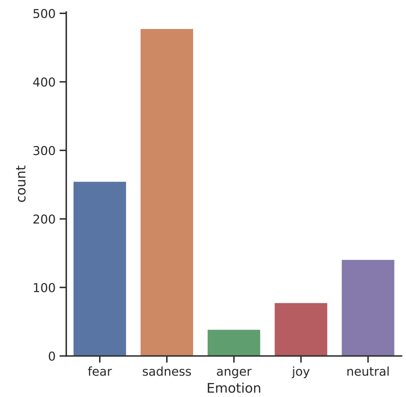

Fig. 5. Emotion classification results from the BERT model.

F. Emotion Classification of SET Responses

The emotional distribution of 991 valid responses (after removing NA values) in the open-ended text SET question is shown in Figure 5. Fear and sadness emotions are prevalent among these gang-involved youth and young adults. Since sadness is the most frequent emotion label and fear is the second most frequent, the model suggests that it is important for GRYD to further investigate strategies and interventions targeting sadness and fear. However, these conclusions are influenced by the model's accuracy on the sadness class, and may vary if accuracy were increased for other classes.

V. Problem and Strategy Classification

A. Text classification for Problem and Strategy

The "problem" column within the SET dataset consists of descriptions of different problems (social, health, financial, etc.) that clients are currently experiencing. The "strategy" column in the dataset consists of a number of workshops/services received participants when receiving GRYD FCM services. Classification of these problems and strategies helps us better evaluate the effectiveness of the GRYD FCM Program. The method we used to classify is the k-means clustering algorithm. We first prepare the input for the algorithm using `TF-IDF` vectorizer, which converts the raw text data to a matrix of feature terms. Then, the optimal numbers of clusters k are determined to be five (for problems) and five (for strategies) using the elbow method.

After clustering, looking at the words in cluster centers allows us to classify problems into five categories: (1) emotional/behavioral issues, (2) school-related issues, (3) probation issues, (4) employment issues, and (5) substance abuse issues. We also classified strategies into five categories: (1) mentoring/support, (2) developmental asset, (3) employment counseling, (4) job/school readiness, and (5) anger management/life skills.

B. Visualization of Problem and Strategy Clustering

We take advantage of PCA and t-SNE techniques for dimension reduction and visualizing clustering results. We first

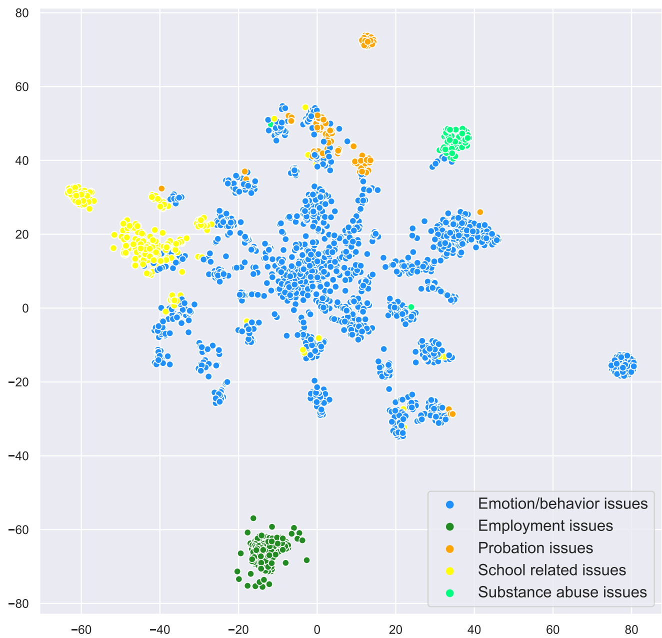

Fig. 6. Text clustering for problems.

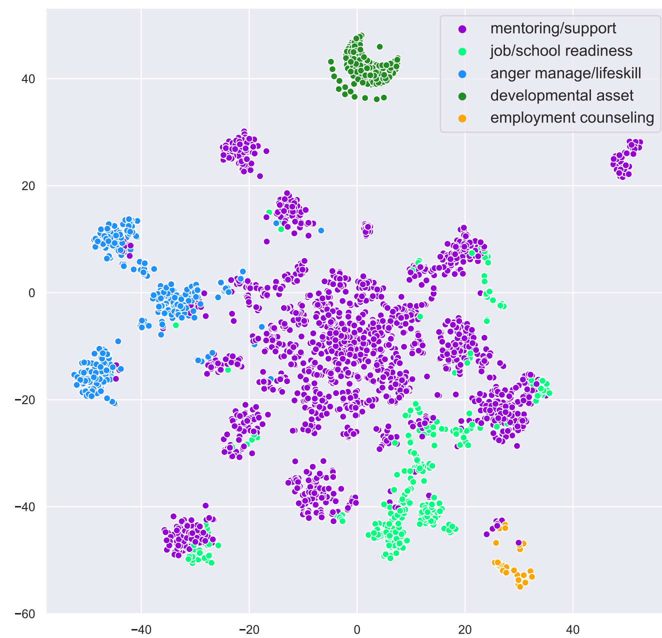

Fig. 7. Text clustering for strategies.

reduce the dimension of text data to 30 with PCA, allowing us to extract main features of the text and filter out irrelevant noise. We then apply the t-SNE visualization algorithm to create a two-dimensional map from our PCA-reduced text data. Figures 6 and 7 show the results of visualizing the problem and strategy clusters, respectively. In both cases, the majority of clusters seem qualitatively well-separated from one another, suggesting k-means clustering is relatively successful on our dataset. However, a limitation is that reducing our dataset to two dimensions eliminates a great deal of the structure that may exist in higher dimensions.

VI. Statistical Analysis of Risk Scores

Risk scores, introduced in Section III, represent an individual's tendency to take risks and engage delinquent behaviors. A high risk score indicates a high possibility of delinquency. We explore risk score distribution based on SET questionnaire data related to problems, strategies, and emotions.

A. Positive Activities

One of the free-form questions in the SET dataset asks respondents to "Describe positive activities." We pre-process the raw data by manually grouping the answers into 12 categories: sports, art, church, job, study, program, self-improve, volunteer, therapy, family, entertainment and other.

The boxplot in Figure 8 shows how risk scores are distributed for each of these activity categories. We observe that the entertainment category (which may include activities like movies, video games, etc.) has the highest median risk score among the categories, while categories like family, job, program, and church tend to have relatively lower risk score. We hypothesize that activities that involve greater commitment and reliability, such as jobs, may contribute to lower risk scores. Programs that emphasize teamwork and responsibility may also contribute significantly to lower risk.

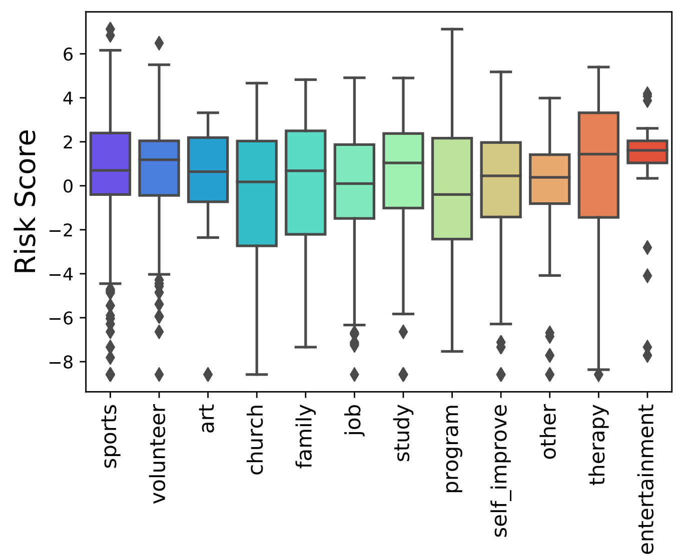

Fig. 8. Boxplot of risk score by activity.

B. Problems and Strategies

The boxplot in Figure 9 shows how risk scores are distributed by each category problem participants face. We observe that participants who are experiencing probation issues tend to have higher risk score than others, though other categories such as emotion/behavior issues exhibit a number of outliers with large risk scores.

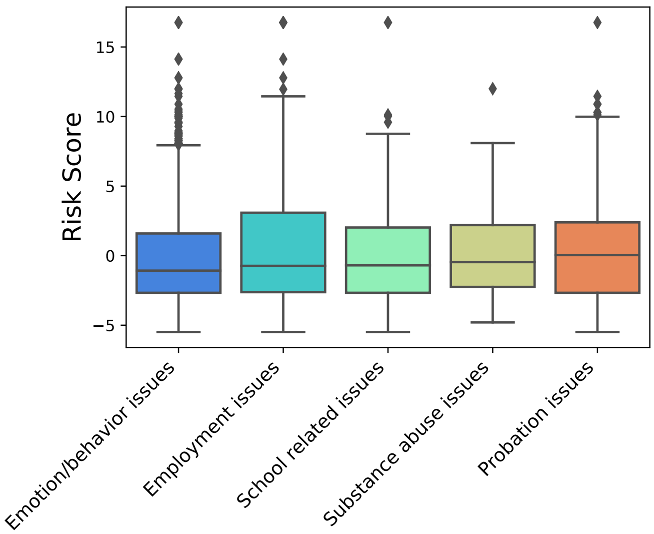

Fig. 9. Boxplot of risk score by problem.

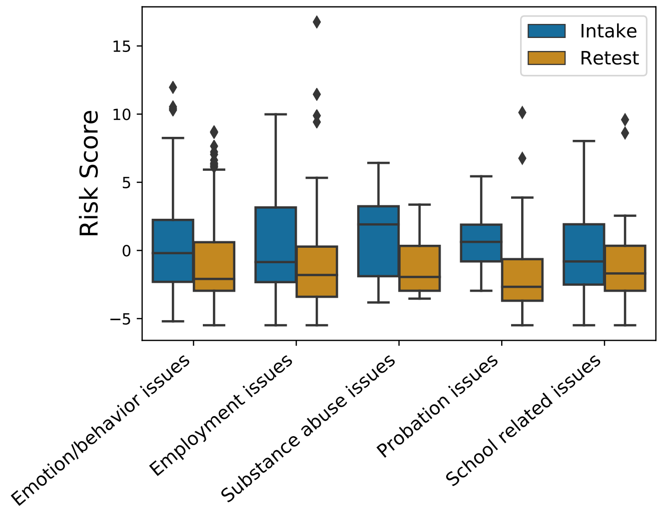

Fig. 11. Boxplot of change in risk score by problem for SET-Intake and SET-Retest.

The boxplot in Figure 10 shows how risk scores are distributed by each type of strategy. We can see that young adults who are provided with employment counseling services tend to correlate with higher risk score than others. Other categories such as mentoring/support exhibit a number of outliers with large risk scores.

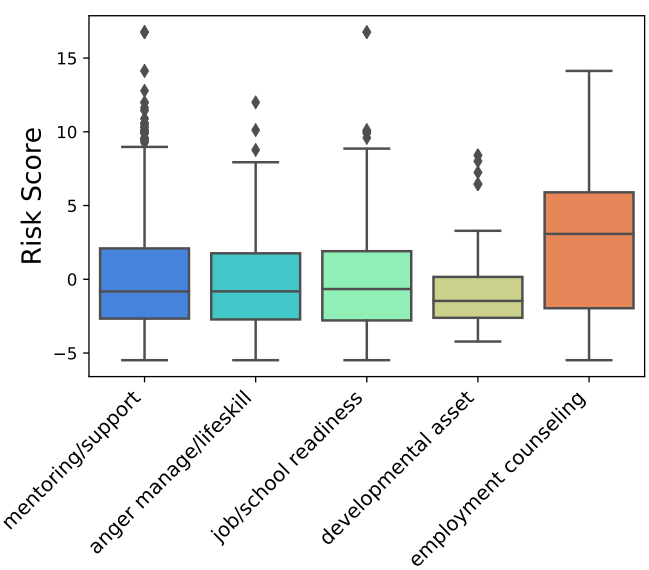

Fig. 10. Boxplot of risk score by strategy.

Within the SET data, some individuals are tested more than once, at different points in time. Accordingly, we consider a subset of the SET data corresponding to SET-Intake and SET-Retest pairs of results. Using this subset of the data, the boxplot in Figure 11 shows that the median risk scores of young adults for all five classes of problems decrease. Most notably, participants who had substance abuse issues and probation issues tend to have significant drops in their median risk score.

Similarly, the boxplot in Figure 12 shows that the median risk scores of young adults for all five classes of strategies decrease. In particular, young adults who engage in anger management/life skills classes tend to have substantial reductions in their risk score while those who receive developmental asset have the least reduction, though this plot suggests that any type of intervention by GRYD FCM providers is, on average, effective.

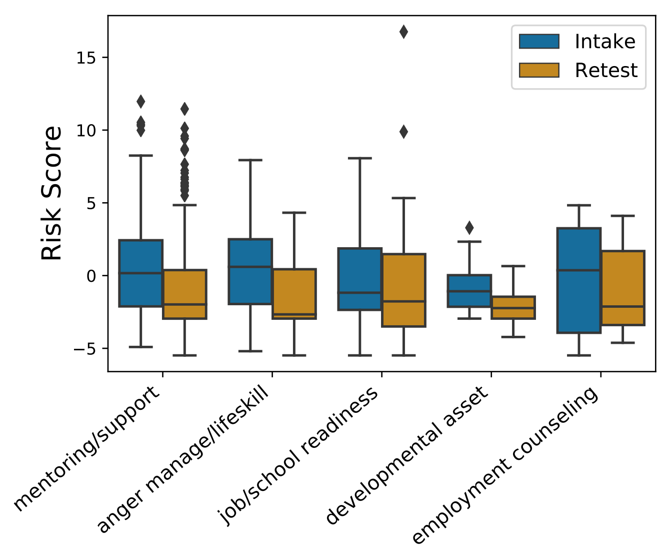

Fig. 12. Boxplot of change in risk score due to strategy for SET-Intake and SET-Retest.

C. *Significance of Changes in Risk Score by Emotion*

Using the SET-intake and SET-Retest pairings of data as in the previous subsection, we explore how individuals' risk scores change over time according to their emotion classification results obtained from the BERT model in Section IV. To assess whether statistically significant changes occurred among the various sub-groups, we conduct one-sided paired t-tests on the emotion classes. See Table II. Based on the p values from the t-tests, there is not enough evidence to suggest that participants exhibiting "anger" and "joy" have significant drops in risk score from SET-Intake to SET-Retest. However, we are able to conclude that participants exhibiting other emotions do have a significant drop in risk score. This suggests two interesting conclusions: (1) GRYD provider interventions are effective in reducing risk score for most subpopulations, when partitioning the population by emotion label, and (2) GRYD may need to invest more in addressing the specialized needs of those who present anger as a regular feature in text responses.

D. *Significance of Changes in Risk Score by Problem Type*

Similarly, we performed one-sided paired t-tests to see whether the drop in risk score (μ_d) for participants with a specific kind of problem is significant following from SET-Intake to SET-Retest. The null hypothesis for the test is

$$H_0 : \mu_d = 0$$

i.e., that there is no difference in risk score from SET-Intake to SET-Retest for participants facing a particular problem. The alternative hypothesis is

$$H_a : \mu_d < 0$$

i.e., that there is a significant decrease in risk score from intake to retest for participants facing a particular problem.

Performing these tests reveals that participants who experienced **emotional/behavioral issues, school-related issues, probation issues, and Substance abuse issues** have significant drops in risk score (see Table II). However, we cannot reject the null hypothesis for individuals who had employment issues. Hence we cannot conclude that the drop in risk score is significant.

E. *Significance of Changes in Risk Score by Strategy Type*

We again applied one-sided paired t-tests to analyze the significance of reductions in risk scores for participants who were treated with various strategies offered by the GRYD FCM Program. From these tests (see Table II), we conclude that strategies in the **mentoring/support, developmental Asset,** and **anger management/life skill** are effective since they all have p-values less than 0.05. However, strategies focusing on providing Employment counseling and Job/school readiness do not produce sufficient evidence to reject the null hypothesis. The drop in risk score for these strategies are not statistically significant.

VII. Conclusions

We explored how risk scores of GRYD FCM participants who completed the SET questionnaire depend on their answers to several free-form text questions. We applied NLP techniques for labeling the emotions conveyed in these responses. We also studied responses related to participants' other (non-gang) activities, their problems, and the GRYD FCM intervention strategies they were exposed to. While these analyses revealed interesting insights on their own, we further considered how participants' risk scores changed over time, breaking down the population by several different factors. We found statistically significant changes for many—but not all—subsets of the population. These results suggest strengths of existing GRYD FCM interventions as well as areas where GRYD FCM can further investigate. In the future, we are interested in applying transfer learning techniques like BERT to broader sets of gang-related data, such as criminal records and other questionnaires. We are also interested in exploring topic modeling approaches for text classification and understanding, particularly for short texts, see e.g. [54].

TABLE II
p-VALUES FOR PAIRED t-TESTS

Emotion	p-value	t-statistic	Sample size
Fear	0.017	2.4	69
Sadness	0.000	4.8	141
Anger	0.285	1.1	10
Joy	0.12	1.63	23
Neutral	0.035	2.2	35
Problem	**p-value**	**t-statistic**	**Sample size**
Emotional/behavioral issues	0.000	7.1	189
School-related issues	0.004	2.1	28
Probation issues	0.003	3.2	32
Employment issues	0.127	1.6	51
Substance abuse issues	0.019	2.7	12
Strategy	**p-value**	**t-statistic**	**Sample size**
Mentoring/support	0.000	6.6	195
Developmental asset	0.000	4.5	20
Employment counseling	0.491	0.7	11
Job/school readiness	0.329	1.0	36
Anger management/life skill	0.000	4.9	50

Acknowledgments

We adapted Python notebooks on machine learning by Lukas Garbas[3] for training models on our chosen datasets. This work began as a UCLA CAM Summer REU project.

The SET was created by the City of Los Angeles Mayor's Office of Gang Reduction and Youth Development and is the copyright of the City of Los Angeles and has been used with permission. Any opinions, findings, conclusions or recommendations expressed in this study, however, are those of the authors and do not necessarily reflect the views of the GRYD Office. This research was funded by the City of Los Angeles contract number C-132202 ("GRYD Research and Evaluation").

[3]https://github.com/lukasgarbas/nlp-text-emotion

REFERENCES

[1] W. B. Miller, *Violence by youth gangs and youth groups as a crime problem in major American cities*. National Institute for Juvenile Justice and Delinquency Prevention, 1977.

[2] J. M. Hagedorn and P. Macon, *People and Folks. Gangs, Crime and the Underclass in a Rustbelt City*. ERIC, 1988.

[3] C. L. Maxson, *Street gangs and drug sales in two suburban cities*. US Department of Justice, 1995.

[4] J. C. Howell, A. Egley Jr., G. E. Tita, and E. Griffiths, "US gang problem trends and seriousness, 1996–2009," *National Gang Center Bulletin*, vol. 6, 2011.

[5] C. R. Huff, "Youth gangs and public policy," *Crime & Delinquency*, vol. 35, no. 4, pp. 524–537, 1989.

[6] M. Stodolska, J. C. Acevedo, and K. J. Shinew, "Gangs of Chicago: Perceptions of crime and its effect on the recreation behavior of Latino residents in urban communities," *Leisure Sciences*, vol. 31, no. 5, pp. 466–482, 2009.

[7] T. J. Taylor, D. Peterson, F.-A. Esbensen, and A. Freng, "Gang membership as a risk factor for adolescent violent victimization," *Journal of Research in Crime and Delinquency*, vol. 44, no. 4, pp. 351–380, 2007.

[8] A. Dıaz-Cayeros, B. Magaloni, A. Matanock, and V. Romero, "Living in fear: Mapping the social embeddedness of drug gangs and violence in Mexico," *Manuscript: University of California at San Diego*, vol. 90, pp. 91–98, 2011.

[9] Q. C. Thurman, A. L. Giacomazzi, M. D. Reisig, and D. G. Mueller, "Community-based gang prevention and intervention: An evaluation of the neutral zone," *Crime & Delinquency*, vol. 42, no. 2, pp. 279–295, 1996.

[10] J. Winterdyk and R. Ruddell, "Managing prison gangs: Results from a survey of U.S. prison systems," *Journal of Criminal Justice*, vol. 38, no. 4, pp. 730–736, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0047235210001029

[11] M. Cahill and D. Hayeslip, "Findings from the evaluation of OJJDP's Gang Reduction Program," *Office of Juvenile Justice and Delinquency Prevention Juvenile Justice Bulletin*, 2010.

[12] I. A. Spergel, *Reducing youth gang violence: The little village gang project in Chicago*. Rowman Altamira, 2007.

[13] G. D. Gottfredson and D. C. Gottfredson, "Gang problems and gang programs in a national sample of schools," 2001.

[14] A. Tremblay, D. Herz, R. Zachery, and M. Kraus, "The Los Angeles Mayor's Office of Gang Reduction and Youth Development comprehensive strategy," 2020, GRYD Research Brief No. 1.

[15] M. Cahill, J. Jannetta, E. Tiry, S. Lowry, M. Becker-Cohen, E. Paddock, and M. Serakos, "Evaluation of the Los Angeles Gang Reduction and Youth Development program: year 4 evaluation report," 2015.

[16] R. Likert, "A technique for the measurement of attitudes." *Archives of psychology*, vol. 22, no. 140, pp. 5–55, 1932.

[17] G. G. Gaes, S. Wallace, E. Gilman, J. Klein-Saffran, and S. Suppa, "The influence of prison gang affiliation on violence and other prison misconduct," *The Prison Journal*, vol. 82, no. 3, pp. 359–385, 2002.

[18] V. Egan and M. Beadman, "Personality and gang embeddedness," *Personality and Individual Differences*, vol. 51, no. 6, pp. 748–753, 2011.

[19] J. Hagan, "The social embeddedness of crime and unemployment," *Criminology*, vol. 31, no. 4, pp. 465–491, 1993.

[20] D. C. Pyrooz, G. Sweeten, and A. R. Piquero, "Continuity and change in gang membership and gang embeddedness," *Journal of Research in Crime and Delinquency*, vol. 50, no. 2, pp. 239–271, 2013.

[21] R. Vargas, "Criminal group embeddedness and the adverse effects of arresting a gang's leader: A comparative case study," *Criminology*, vol. 52, no. 2, pp. 143–168, 2014.

[22] G. Sweeten, D. C. Pyrooz, and A. R. Piquero, "Disengaging from gangs and desistance from crime," *Justice Quarterly*, vol. 30, no. 3, pp. 469–500, 2013.

[23] F. Gibou, D. Hyde, and R. Fedkiw, "Sharp interface approaches and deep learning techniques for multiphase flows," *Journal of Computational Physics*, vol. 380, no. 442–463, 2019.

[24] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams engineering journal*, vol. 5, no. 4, pp. 1093–1113, 2014.

[25] R. K. Bakshi, N. Kaur, R. Kaur, and G. Kaur, "Opinion mining and sentiment analysis," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 2016, pp. 452–455.

[26] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[27] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. J. Passonneau, "Sentiment analysis of Twitter data," in *Proceedings of the workshop on language in social media (LSM 2011)*, 2011, pp. 30–38.

[28] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep convolution neural networks for Twitter sentiment analysis," *IEEE Access*, vol. 6, pp. 23 253–23 260, 2018.

[29] Q. Zhou and H. Wu, "NLP at IEST 2018: BiLSTM-attention and LSTM-attention via soft voting in emotion classification," in *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2018, pp. 189–194.

[30] J. Yu, L. Marujo, J. Jiang, P. Karuturi, and W. Brendel, "Improving multi-label emotion classification via sentiment classification with dual attention transfer network," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 1097–1102.

[31] Y. Lai, L. Zhang, D. Han, R. Zhou, and G. Wang, "Fine-grained emotion classification of Chinese microblogs based on graph convolution networks," *World Wide Web*, vol. 23, no. 5, pp. 2771–2787, 2020.

[32] A. Seyeditabari, N. Tabari, and W. Zadrozny, "Emotion detection in text: a review," *arXiv preprint arXiv:1806.00674*, 2018.

[33] L. A. M. Bostan and R. Klinger, "An analysis of annotated corpora for emotion classification in text," in *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics, 2018, pp. 2104–2119. [Online]. Available: http://aclweb.org/anthology/C18-1179

[34] V. Liu, C. Banea, and R. Mihalcea, "Grounded emotions," in *International Conference on Affective Computing and Intelligent Interaction (ACII 2017)*, San Antonio, Texas, 2017.

[35] C. Strapparava and R. Mihalcea, "Semeval-2007 task 14: Affective text," in *Proceedings of the 4th International Workshop on Semantic Evaluations*. Association for Computational Linguistics, 2007, pp. 70–74.

[36] H. Schuff, J. Barnes, J. Mohme, S. Pad'o, and R. Klinger, "Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus," in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2017, pp. 13–23.

[37] D. Preoţiuc-Pietro, H. A. Schwartz, G. Park, J. Eichstaedt, M. Kern, L. Ungar, and E. Shulman, "Modelling valence and arousal in Facebook posts," in *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2016, pp. 9–15.

[38] S. Buechel and U. Hahn, "EMOBANK: Studying the impact of annotation perspective and representation format on dimensional emotion analysis," *EACL 2017*, p. 578, 2017.

[39] Y. Li, H. Su, X. Shen, W. Li, Z. Cao, and S. Niu, "DailyDialog: A manually labelled multi-turn dialogue dataset," *arXiv preprint arXiv:1710.03957*, 2017.

[40] C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 2005, pp. 579–586.

[41] D. Ghazi, D. Inkpen, and S. Szpakowicz, "Detecting emotion stimuli in emotion-bearing sentences." in *CICLing (2)*, 2015, pp. 152–165.

[42] S. Aman and S. Szpakowicz, "Identifying expressions of emotion in text," in *Text, speech and dialogue*. Springer, 2007, pp. 196–205.

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[44] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc., 2009.

[45] Y. Kim, "Convolutional neural networks for sentence classification," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

[46] Y. Zhang and B. C. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," 2016.

[47] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin, "Advances in pre-training distributed word representations," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[48] M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," *Signal Processing, IEEE Transactions on*, vol. 45, pp. 2673 – 2681, 12 1997.

[49] Y. B. Dzmitry Bahdanau, Kyunghyun Cho, "Neural machine translation by jointly learning to align and translate," 2014.

[50] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[51] M. Sundermeyer, R. Schlüter, and H. Ney, "LSTM neural networks for language modeling," in *Thirteenth annual conference of the international speech communication association*, 2012.

[52] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014.

[53] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2019.

[54] J. Qiang, Z. Qian, Y. Li, Y. Yuan, and X. Wu, "Short text topic modeling techniques, applications, and performance: a survey," *IEEE Transactions on Knowledge and Data Engineering*, 2020.