International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 5, pp: 786-789 www.ijaem.net ISSN: 2395-5252

Review: Natural Language Processing and Speech Signal for Mental Health Research and Care Application

Dr.S.N.Kakarwal, Pradip M. Paithane, Sushant S.Khedikar
[1,]*Professor, CSE Department, PES College, Aurangabad, Maharashtra, India*
[2]*Assistant Professor, Computer Engineering Department, VPKBIET, Baramati, Maharashtra,India*
[3]*Assistant Professor, Electronics Department, PES College, Aurangabad, Maharashtra, India*
Corresponding author; Dr.S.N.Kakarwal

Date of Submission: 30-08-2020 Date of Acceptance: 11-09-2020

ABSTRACT: Natural language processing (NLP) techniques can be used to make inferences aboutpeoples' mental states from what they write on Facebook, Twitter and other social media. Speech signal processing is being increasingly explored in health domains given both its centrality as a behavioral cue and the promise of robust, automated analysis of data at scale. We discuss general issues in health-related speech research. Focus two health applications we have undertaken: autism spectrum disorder (unsupervised) and addiction counseling (supervised). Methods range from deep supervised learning to knowledge-based signal processing of highly subjective constructs of psychological states and traits. Atypical Prosody is the way in which something is said–the intonation and rhythm of speech. A sole aspect of the research has been to model both health care provider and patient behaviors jointly in clinical encounters, wherein any individual behavior cannot be measured in isolation given the inherent mutual impact.

KEYWORDS: Behavioral signal processing (BSP), deep learning, knowledge-inspired features, addiction counseling, autism spectrum disorder, atypical prosody

I. INTRODUCTION

Mental illness is one of the most pressing public health issues. Speech contains rich information beyond the spoken words that is critical for communicating intent as well as paralinguistic information such as emotional state. A person's voice is also in many ways idiosyncratic, but certain components provide a listener indication of speaker attributes like age and gender. Moreover, a person's voice changes subject to transient factors such as having a cold or being intoxicated. There is now significant interest in quantifying all aspects of voice, particularly those related to mental health as the complex processes underlying speech are often affected by neurological and motor disorders. While more established speech tasks like automatic speech recognition have vast data reserves, available data is often quite limited for emotional and medical conditions. Computational models of speech affect have been explored for little more than a decade (e.g., [1]), and the focus is only now shifting to the challenge of "in the wild" interfaces [2].

The exponential growth in speech processing applications for various mental and physiological health conditions is an even more recent phenomenon; examples include depression [3], Parkinson's disease [4], and autism spectrum disorder (ASD) [5]. Multi-modal systems will eventually become a focus, and work is concurrently being done in physiological [6] and facial expression analysis [7]. We cast these problems through the lens of Behavioral Signal Processing (BSP; [8]), which encompasses the mapping from low-level behavioral signals to high-level human behavioral constructs in order to support and augment a domain expert's decision-making. Each application domain requires a unique set of considerations in the problem formulation stage. Two different computational approaches are studied for the modeling, one in rating therapist empathy in drug addiction counseling sessions and the other in modeling speech prosody in autism spectrum disorder.

II. METHODOLOGICAL CHALLENGES AND CONCERNS

Human behavior is highly complex and variable. For a single person, moment-to-moment changes internally and externally affect an acquired signal. Heterogeneity of human behavior and partial, noisy acquisition of behavioral signals are two primary challenges in developing reliable BSP systems. Another issue is the subjectivity of human ratings, for which several works have proposed

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 5, pp: 786-789 www.ijaem.net ISSN: 2395-5252

solutions [11], [12]. Other issues relate to problem formulation, such as whether to rely on knowledge based feature design, brute force feature over generation, or a black-box feature generation approach; the appropriate method depends on the specific application, the available data, and the degree of interpretability required. A related issue is system generalization.

III. EMPATHY IN ADDICTION COUNSELLING

A. Empathy Modeling: Background

Counselor empathy is of special interest in addiction counseling as it relates to the counselor's ability to establish a rapport with their clients as well as successful outcome measures [13]. In motivational interviewing, a client-based psychotherapy, empathy is defined as "the extent to which the therapist understands and/or makes an effort to grasp the client's perspective" [14]. Recently there have been a number of studies focused on characterizing counselor empathy in addiction counseling using lexical, prosody, speech rate, and vocal entrainment cues [15], [16], [17].

The data were manually transcribed and segmented at the utterance level. Each utterance received a local behavior label from the motivational interviewing skill code (MISC) manual and the full session received a global empathy rating according to the to the motivational interviewing treatment integrity (MITI) manual [14].

B. Empathy Modeling: Approach

Employ a deep learning approach to first map from raw features to counselor- and client-utterance-level behavioral categories; then from utterance-level behaviors. Use an encoder-decoder network where the encoder learns the local-level behavioral encoding and the decoder does the global-level behavior prediction. The local behaviors in each turn, of the ith session, are represented by a L-dimensional, k-hot vector, Yit, for the k utterance level behaviors which occur in that turn. The L behaviors which are considered are simple and complex reflections, open and closed questions, facilitation, giving information, other counselor behavior, and client behavior. The global behavior, zi, is 1 for 'high empathy' and 0 for 'low empathy'.

The encoder takes turn level word embedding vectors as input. This feature vector is the average of the word embedding vectors belonging to the words that occur in the turn, i.e.

$$X_t = \frac{1}{|W_t|} \sum_{w \in W_t} V_t$$

where Wt is the set of words in turn t and Vw is the embedding vector for word, w. The sequence of turn feature vectors in turn t, Xti , is passed to a dense feed-forward layer which performs a feature transformation on the turn embedding feature vectors, resulting in Xti . Next, the transformed vector is input to a recurrent layer which is used to learn context across turns. Subsequently, the output of the recurrent layer, Hti , is input to a feed-forward layer with sigmoid activation which performs a multi-label prediction, Yti , of the behaviors in turn t [1][4][6].

The decoder, maps from the local behavioral encoding, Yti, to the predicted global behavior ,Zi. To achieve this it receives the local behavior encoding as input from the encoder, performs a mean pooling across turns, resulting in Yi, which is passed to a feed-forward layer with sigmoid activation to predict the global behavior[4].

The word embedding vectors are multi-dimensional and were trained using the word2vec software [2]. The deep neural network models were trained using the Keras deep learning library with Theano back-end [16].

IV. AUTISM SPECTRUM DISORDER

Autism spectrum disorder (ASD) is a neuro developmental disorder defined by impairments in social communication and reciprocity, as well as restricted, repetitive behavioral patterns and interests. ASD certainly has neurological and genetic underpinnings; diagnosis is currently based on behavioral observation from expert human clinicians.

A. ASD: Approach

Reliable measure of atypical prosody—is what made the formulation difficult; clinicians could use a better measure of atypical prosody, but no reliable reference with which to train a system. Current approach is to create a bottom-up definition of various facets of atypical prosody, and then validate its utility through analysis with related measures including ASD diagnosis and subjective ratings of prosodic atypicality [2]. It's important to clarify why the existing reference labels that are in common use in the limited amount of acoustic-prosodic research in ASD are insufficient. The most common approach sets diagnosis as the target variable, assuming that a unique "autistic prosody"

exists; but the various prosodic atypicalities will not be universal in or unique to ASD. Alternatively, subjective human ratings are hindered by poor inter-rater reliability. Proposed method follows a bottom up, rule-based definition of atypical prosody [14].

B. ASD: Proposed Features

Motivated by qualitative depictions of what atypical prosody entails as well as linguistic research into acoustic correlates of perception, discuss various novel features of atypical prosody. These features can be categorized in the domains of intonation, rate and voice quality. Intonation features include micro-prosodic pitch and intensity contour features as well as macro-prosodic, utterance-level modeling of dynamics [10]. Voice quality features are motivated by the global perceptions associated with autism such as harshness and nasality; one of the most informative features has proven to be jitter. Proposed a novel feature which measures the coordination of multiple prosodic streams utilizing pairwise correlation; individuals with ASD are hypothesized to have poor timing and coordination.

V. CONCLUSION

BSP is an extremely promising area of study, particularly with the high-impact possible in health domains. But each problem is unique, and challenges exist in developing a practical problem formulation, robust signal acquisition and signal processing, and useful behavioral informatics. In the first approach, supervised approach can produce a system with real world applicability through deep learning on acoustics using modeled the empathy.
In the second, an unsupervised approach is used atypical prosody in autism. In this, rule-based, bottom-up methods can be used to create a novel, objective measure of a behavioral construct which is unreliable for human raters.

In the future, introduced massive-scale, longitudinal analysis of speech which can predict and track the progression of disorders like Parkinson's disease.

REFERENCES

[1]. Chul Min Lee, Shrikanth Narayanan, and Roberto Pieraccini,"Recognition of negative emotions from the speech signal," in Automatic Speech Recognition and Understanding, 2001. ASRU' 01. IEEE Workshop on. IEEE, 2001, pp. 240–243.

[2]. Jun Deng, Zixing Zhang, Florian Eyben, and Bj¨orn Schuller, "Autoencoder-based unsupervised domain adaptation for speech emotion recognition," IEEE Signal Processing Letters, vol. 21, no.9, pp. 1068–1072, 2014.

[3]. Nicholas Cummins, Vidhyasaharan Sethu and Jarek Krajewski, "Analysis of acoustic space variability in speech affected by depression," Speech Communication, vol. 75, pp. 27–49, 2015.

[4]. Juan Rafael,Elkyn Alexander Belalcazar-Bolanos,Juli´an David Arias-Londo˜no and Elmar N¨oth, "Characterization methods for the detection of multiple voice disorders: Neurological, functional, and laryngeal diseases," IEEE journal of biomedical and health informatics, vol. 19, no. 6, pp.1820–1828, 2015.

[5]. Daniel Bone, Matthew P Black, Chi-Chun Lee and Shrikanth Narayanan, "The Psychologist as an Interlocutor in Autism Spectrum Disorder Assessment: Insights from a Study of Spontaneous Prosody," Journal of Speech, Language, and Hearing Research, vol. 57, pp. 1162–1177, 2014.

[6]. Tanaya Guha, Sungbok Lee and Shrikanth S Narayanan,"On quantifying facial expression-related atypicality of children with autism spectrum disorder," in 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2015, pp. 803–807.

[7]. S. Narayanan and P. G. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," Proceedings of the IEEE, vol. PP, no. 99, pp. 1–31, 2013.

[8]. Chi-Chun Lee, Athanasios Katsamanis and Shrikanth S Narayanan, "Computing vocal entrainment: A signal derived pca-based quantification scheme with application to affect analysis in married couple interactions," Computer Speech & Language, vol. 28, no. 2, pp. 518–539, 2014.

[9]. Daniel Bone, Somer Bishop,Sungbok Lee, and Shrikanth Narayanan, "Acoustic-prosodic and turn-taking features in interactions with children with neurodevelopmental disorders," in Proceedings of Inter speech, 2016, pp. 1185–1189.

[10]. Rahul Gupta, Kartik Audhkhasi and Shrikanth Narayanan,"Modeling multiple time series annotations based on ground truth inference and distortion," IEEE on Affective Computing, vol. PP, 2016.

[11]. Robert Elliott, Arthur C Bohart, Jeanne C Watson, and Leslie S Greenberg, "Empathy.," Psychotherapy, vol. 48, no. 1, pp. 43, 2011.

[12]. Bo Xiao, Panayiotis G Georgiou, and S Narayanan, "Analyzing the language of therapist empathy in motivational interview based psychotherapy," in Asia-Pacific Signal and Information Processing Association, 2012, pp. 1–4.

[13]. Bo Xiao, Panayiotis G Georgiou, Zac E Imel, David C Atkins,and Shrikanth Narayanan, "Modeling therapist empathy and vocal entrainment in drug addiction counseling.," in INTERSPEECH, 2013, pp. 2861–2865.

[14]. Bo Xiao, Zac E Imel, David C Atkins, Panayiotis G Georgiou, and Shrikanth S Narayanan, "Analyzing speech rate entrainment and its relation to therapist empathy in drug addiction counseling," in

[15]. Sixteenth Annual Conference of the International Speech Communication Association, 2015.

[16]. James Gibson,Atkins,Panayiotis Georgiou, and Shrikanth Narayanan, "A deep learning approachto modeling empathy in addiction counseling," Commitment,vol. 111, pp. 21, 2016.

[17]. J. Baio, "Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network,11 sites, united states, 2010.," Morbidity and mortality weekly report.Surveillance summaries (Washington, DC: 2002), vol. 63, no.2, pp. 1, 2014.

[18]. Daniel Bone, Somer Bishop, Matthew P Black and Shrikanth S Narayanan, "Use of machine learning to improve autism screening and diagnostic instruments:effectiveness, efficiency, and multi-instrument fusion," Journal of Child Psychology and Psychiatry, 2016.

[19]. Manoj Kumar, Daniel Bone, Nikolaos Malandrakis and Somer Bishop, "Objective language feature analysis in children with neurodevelopmental disorders during autism assessment," 2016, pp. 2721–2725.

[20]. Daniel Bone, Matthew S Goodwin, Matthew P Black, Chi-Chun Lee, Kartik Audhkhasi, and Shrikanth Narayanan, "Applying machine learning to facilitate autism diagnostics: Pitfalls and promises,"Journal of autism and developmental disorders, vol. 45, no. 5, pp.1121–1136, 2015.

[21]. C. Lord, S. Risi, L. Lambrecht, E. Cook, B. Leventhal, P. DiLavore,A. Pickles, and M. Rutter, "The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism," Journal of Autism and Developmental Disorders, vol. 30, pp. 205–223, 2000.

[22]. Matthew P. Black, Daniel Bone, Marian E. Williams, Phillip Gorrindo,Pat Levitt, and Shrikanth S. Narayanan, "The USC CARE Corpus: Child-Psychologist Interactions of Children with Autism Spectrum Disorders," in Proceedings of Interspeech, 2011.

[23]. Daniel Bone, Matthew P Black, Anil Ramakrishna, Ruth Grossman,and Shrikanth Narayanan, "Acoustic-prosodic correlates of awkwardprosody in story retellings from adolescents with autism," inINTERSPEECH, 2015.

[24]. Daniel Bone, Chi-Chun Lee, Theodora Chaspari, Matthew Black,MarianWilliams, Sungbok Lee, Pat Levitt, and Shrikanth Narayanan,"Acoustic-prosodic, turn-taking, and language cues in childpsychologist interactions for varying social demand," in INTERSPEECH,2013.

[25]. Christiane AM Baltaxe, "Use of contrastive stress in normal, aphasic,and autistic children," Journal of Speech, Language, and Hearing Research, vol. 27, no. 1, pp. 97–105, 1984.