November 2019

Robust ADHD testing by applying clustering techniques to survey responses or speech data

Lingyi Zhang

Cliff Chin Ngai Sze

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Robust ADHD testing by applying clustering techniques to survey responses or speech data

ABSTRACT

Existing tests for attention deficit hyperactivity disorder (ADHD) may exhibit some bias. Also, these tests require filling in a survey with subjective responses, which can lead to misdiagnosis. The techniques described herein reduce bias in ADHD tests by seeking clusters in test-parameter space conditioned on certain characteristics of a person. Clustering is performed using machine learning techniques. With user permission, speech data is obtained via one or more devices such as a phone, smart speaker, etc. an is used to make objective diagnoses of ADHD.

KEYWORDS

- Attention deficit hyperactivity disorder (ADHD)

- Test bias

- Clustering

- Machine learning

BACKGROUND

There are several parameters that a person is tested for in order to arrive at a diagnosis of ADHD. For example, the Vanderbilt assessment scale (VAS) has fifty-five questions that a person (or their caretaker) answers in order for a doctor to arrive at an ADHD diagnosis. Other survey-based diagnostic tools for ADHD include child attention profile (CAP), behavior assessment system for children (BASC), child behavior checklist (CBCL), etc. Example questions from the Vanderbilt assessment scale include: "does the person not pay attention to details or makes careless mistakes with, for example, homework," "does the person have

difficulty keeping attention to what needs to be done," etc., with each question requiring a numeric response on a 0-to-3 scale.

Existing tests for attention deficit hyperactivity disorder (ADHD) may exhibit some bias. For example, children born in August receive a 34% higher rate of ADHD diagnosis than their September-born peers, a phenomenon explained by the relative youth of August-born children when compared to others in their school-year cohort.

Another problem with ADHD testing is that the person (or their caretaker) is required to fill a survey with responses to subjective questions that depend on human recall, such as "does the person not seem to listen when spoken to directly," or "does the person blurt out answers before questions have been completed."

DESCRIPTION

This disclosure presents techniques to remove bias in ADHD testing and to objectively assess certain parameters that are used for the diagnosis of ADHD.

Removing bias in ADHD testing

Per techniques of this disclosure, each question from a diagnostic tool such as VAS, CAP, BASC, CBCL, etc. is treated as a dimension in a multi-dimensional space, such that the set of responses to the diagnostic tool constitutes a feature vector. A machine learning technique, e.g., anomaly detection, is used to cluster the feature vectors. The formation of two or more clusters is indicative of persons falling in two or more categories. The categories can be identified by a medical practitioner as being consistent with ADHD or not (or falling in an in-between spectrum).

Furthermore, with permission from the person to use one or more specific characteristics, clustering can be performed conditioned on selected characteristics of the person, e.g., the

person's birth-month, gender, prematurity of birth, age when the person started walking (or spoke first words), medical history, psychological conditions, developmental milestones (typically in these categories: Social/Emotional, Language/Communication, Cognitive, Movement/Physical Development), or other dimensions of the feature vector. When thus conditioned, biases such as the above-described August-born children receiving a higher rate of ADHD diagnoses disappear, as the ADHD classification of a person takes into account the person's birth-month cohort. Conditioning on selected characteristics of the person moves the threshold-plane that determines ADHD which can result in more accurate diagnoses. The clustering of feature vectors to detect ADHD can also reveal correlations between various questions on the diagnostic tool, thereby enabling the simplification of the survey, e.g., by eliding or combining questions, etc. Since the feature set utilized is much larger than the existing ADHD tests, this approach can generate new test(s) based on the anomaly detection (or other machine learning techniques) as described in the next section.

Objectively assessing persons for ADHD, and the discovery of new ADHD tests

Certain questions on ADHD diagnostic tools are based on the observed verbal behavior of the person. Some examples of such questions are:

- Does the person not seem to listen when spoken to directly?
- Does the person blurt out answers before questions have been completed?
- Does the person interrupt or intrude in on others' conversations and/or activities?

The person (or their caretaker) typically answers such questions on a 0-3 scale.
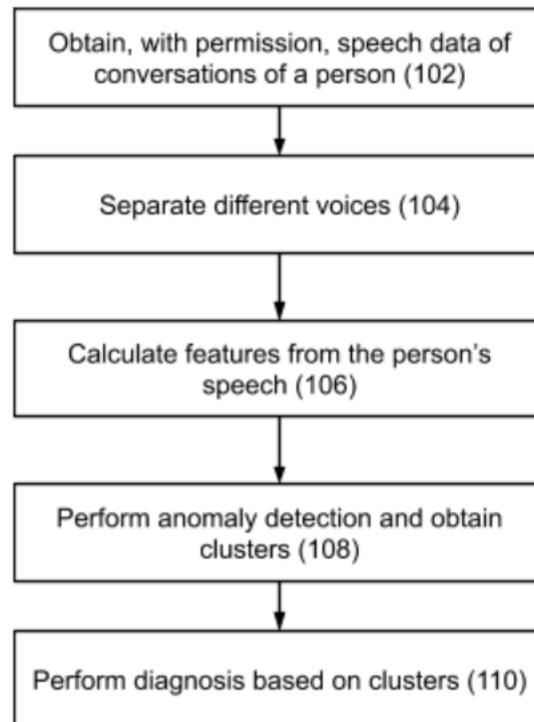
Fig. 1: Performing diagnosis from speech data

Per the techniques of this disclosure, speech data of a person's speech in everyday settings (102) is obtained via a phone, tablet, smart speaker, or other device, with user permission. Such data can potentially be obtained over long intervals and at different times of the day. Different voices in the speech data are separated (104). After isolating the person's voice, features are determined based on the person's speech (106).

With user permission, the features determined based on the person's speech, together with features obtained from other permitting individuals are clustered, e.g., using unsupervised machine learning techniques (108). The features are obtained such that no individually identifiable information is included and are used specifically for unsupervised learning purposes. Users are provided with options to allow or disallow the use of features determined from their speech for clustering. Based on the clustering, new ADHD tests can be identified based on unsupervised machine learning techniques. The clusters thus obtained are then analyzed (110) to

objectively answer questions relating to the user's verbal behavior, such as those listed above, to perform diagnosis. Note that these new features may not be covered by any of the existing ADHD tests.

For example, a virtual assistant application can detect an inordinately long response time to a question posed to the person by someone in the vicinity. When the user permits, new features, e.g., frequency of the user's speech relative to others in a conversation, repetitiveness in a user's speech, etc. can be gleaned from user's speech data and tested for correlation with ADHD. The validity of a new feature as a test for ADHD can be continuously verified and each test can be assigned a confidence score.

The described techniques can be implemented via hardware such as smart speakers, smart displays, phones, tablets, computers, etc. and software, e.g., a virtual assistant application. The described techniques (or portions thereof) are implemented with specific user (and caretaker) permission. Users are provided with options to change their permission settings at any time. Data obtained from the users is utilized only for the specific purposes as described herein. Storage and/or processing of such data is performed in compliance with applicable regulations and guidelines. Users are provided with indications that user data may be used for diagnosis. The data clustering and diagnosis is performed under the guidance and/or control of a medical practitioner.

Further to the descriptions above, a user is provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's voice interaction with a device, a user's speech during conversation with other users, a user's preferences, a user's current location), and if the user is sent content or communications from a server. In addition,

certain data is treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. For example, the user's voice is modified such that it is not identifiable. Further, names, locations, etc. that are mentioned during a conversation are removed. Thus, the user has control over what information is obtained about the user, how that information is used, and what information is provided to the user.

CONCLUSION

The techniques described herein reduce bias in ADHD tests by seeking clusters in test-parameter space conditioned on certain characteristics of a person. Clustering is performed using machine learning techniques. With user permission, speech data is obtained via one or more devices such as a phone, smart speaker, etc. an is used to make objective diagnoses of ADHD.

REFERENCES

[1] Webmd.com, "ADHD tests," https://www.webmd.com/add-adhd/childhood-adhd/adhd-tests-making-assessment#1-2 accessed on Oct. 23, 2019.

[2] Pbs.org, "Could being born in August make an ADHD diagnosis more likely?" https://www.pbs.org/newshour/health/could-being-born-in-august-make-an-adhd-diagnosis-more-likely accessed on Oct. 23, 2019.

[3] National institute for children's health quality, "NICHQ Vanderbilt assessment scales used for diagnosing ADHD," https://www.nichq.org/sites/default/files/resource-file/NICHQ-Vanderbilt-Assessment-Scales.pdf

accessed on Oct. 23, 2019.

[4] Webmd.com, "Mental health and conduct disorder," https://www.webmd.com/mental-health/mental-health-conduct-disorder#1 accessed on Oct. 23, 2019.

[5] Webmd.com, "Anxiety and panic disorders health center," https://www.webmd.com/anxiety-panic/default.htm accessed on Oct. 23, 2019.

[6] Webmd.com, "Depression health center," https://www.webmd.com/depression/default.htm accessed on Oct. 23, 2019.

[7] Webmd.com, "Test your knowledge of stress and anxiety,"

https://www.webmd.com/balance/stress-management/rm-quiz-stress-anxiety accessed on Oct. 23, 2019.

[8] Webmd.com, "Depression overview slideshow,"

https://www.webmd.com/depression/ss/slideshow-depression-overview accessed on Oct. 23, 2019.

[9] Webmd.com, "Health and parenting guide,"

https://www.webmd.com/parenting/guide/default.htm accessed on Oct. 23, 2019.