Turing a la Freud: Test for an Automated Psychiatrist

Guillermo A. Cecchi and Irina Rish
IBM T.J. Watson Research Center
Yorktown Heights, NY 10598

Abstract

We propose to focus the efforts of the Artificial Intelligence community on the problem of diagnosing psychiatric disorders from speech produced during psychiatric interviews. An AI system will be asked to match (or even surpass) the classification accuracy of a human psychiatrist in detecting various mental state disorders, such as schizophrenia, bipolar, depression, anxiety, etc., from the text of an interview. The test can be potentially extended to include not just text, but also audio and video feeds, to enrich the information available to the automated psychiatrist. Moreover, we propose a more challenging version of the test, requiring the AI system to fully imitate a human psychiatrist.

The Game: Consider the following exchange:

[Q] The wind blows because . . .
[A]: Because it’s time to blow.
[Q]: What makes it blow?
[A]: The air.
[Q]: The air?
[A]: The sky.
[Q]: How does the sky make it blow?
[A]: Because it’s high in the air.

It is a classical example of an interview between a psychiatrist and a patient - diagnosed with schizophrenia in this case (McKenna and Oh 2005). An Imitation Game where one participant needs to determine the class as well as various scalar fields of the other, not knowing whether he/she will collaborate or obstruct.

Mental Health: Mental disorders are a cause of suffering for those who experience them, as well as of distress for their families and caregivers. Beyond these ineffable aspects, mental conditions account for a large proportion of health cost in society. The Veterans Affairs estimates 21 million veterans, with a 20% incidence of Post Traumatic Stress Disorder for those returning from the war front. A 2009 report by the National Center on Addiction and Substance Abuse estimates the yearly cost of drug and alcohol abuse in the USA is $460 billion, including care costs and lost work hours. Depression, a condition leading to high morbidity and mortality, and low productivity, is estimated to cost $100 billion in the US and €120 billion in Europe (Sobocki et al. 2006). Adding the cost of other conditions, including obesity and eating disorders considered as a form addiction (Volkow and Wise 2005), estimated recently to be $114 billion (Tsai, Williamson, and Glick 2011), the economic burden of mental disorders approximates the staggering sum of a trillion dollars per year in the US.

The psychiatric interview is the main assessment tool for diagnosis and treatment evaluation. There are several different prescriptions to evaluate specific mental aspects, but in general they involve a semi-structured questionnaire lasting between 30 and 90 minutes, requiring the production of scales upon which the diagnosis is based. For instance, disorders of thought (e.g. psychosis), require the evaluation of specific ways in which the thought process is affected (APA 2013). These include conceptual disorganization and poverty of speech, as well as concomitant emotional components such as anxiety and depression, as expressed in speech (Kay, Fiszbein, and Opler 1987). The example above is just a glimpse.

In the US alone, there are 63.3 million interviews performed every year. With the exception of a handful of isolated studies, the interviews and diagnostic assessments are evaluated without any help of Artificial Intelligence, Natural Language Processing or even run-of-the-mill Machine Learning techniques. Assessment validation studies are performed comparing results across different institutions and countries (Williams et al. 1992), as well as blind studies involving actors (Wallace, Rao, and Haslam 2002).

However, there is a small body of research on the development of analytic tools to capture relevant features present in speech, voice and other readily accessible measures. Focusing on transcribed interviews, (Mota et al. 2012; 2014) present a text-analysis approach based on syntactic graphs and their topological properties, that allows for a very accurate (above 90%) classification of mental disorders such as bipolar disorder and schizophrenia, from relatively short interviews with patients. Similarly, (Bedi et al. 2014) using semantic analysis (B. Elvevag 2007) to study speech alteration effects of psychoactive drugs, demonstrate that automated semantic speech analyses can capture sub-
tle alterations in mental state, accurately discriminating between drugs. There are efforts focusing on acoustic features of speech, known to be particularly affected in certain conditions. This approach was used to accurately discriminate between elderly controls versus the same-age patients with mild cognitive impairment (MCI) and Alzheimer’s disease (Satt et al. 2013). These findings illustrate the potential for automated speech-based approaches to characterize clinically-relevant alterations in mental state, including those occurring in psychiatric illness.

**The Turing a la Freud Test:** We propose the challenge to design a test which should satisfy the following conditions, in (possibly) increasing level of difficulty:

1. **Pass the diagnosis test:** are you schizophrenic? The system should be able to identify whether it is interacting with a patient or a control subject, determine a diagnosis and produce psychometric scales consistent with those produced by a group of trained psychiatrists.

2. **Pass the self-psychiatrist test:** am I a psychiatrist? The previous test does not require the system to emulate a psychiatrist, and doing so should be harder. The system should be able to emulate a psychiatrist, conducting an interview with patients and subjects, so that both groups cannot distinguish it from a trained psychiatrist.

3. **Pass the psychiatrist test:** are you a psychiatrist? The system should be able to identify whether it is interacting with a psychiatrist or a layman. That is, the system should be able to understand that its mental health is being scrutinized.

4. **Pass the self diagnosis test:** am I schizophrenic? The system should be able to emulate the traits that accompany specific mental disorders, so that a trained psychiatrist would diagnose it accordingly.

In an initial stage, the tests would be double-blind and conducted only through transcribed speech or typed text. In later stages additional features such as audio, video and wearable monitors (heart rate, skin conductance, etc.) where appropriate (mostly Test 1).

**Benefits for Psychiatry:** The benefits of a system that can successfully pass Test 1 are numerous and almost self-evident, but we can list the two main ones. Firstly, it would provide for a better understanding of the condition, as the different conditions will be quantified and modeled and add to the current informal definitions. However, it is in the area of prevention that this would have a significant immediate impact. That is, the ability to rapidly and constantly respond and monitor people who are at high risk of developing an outbreak of psychosis, anxiety or depression, leading to accidents, injuries and even suicide, can allow for early intervention and a more rational deployment of limited resources (i.e. psychiatrists and caregivers).

**Benefits for Artificial Intelligence:** Why would AI benefit, beyond the luster of radically changing the way a clinical discipline is practiced? There is increasing consensus in psychiatry that most mental disorders are extreme cases of traits shared by the normal population, in lower degrees and shorter periods of time. A reflection, perhaps, of evolution adaptations, as in depression as a corollary of submission (Stevens and Price 2000). An example, equally cute and terrifying, is the diagnosis of "moderately schizophrenic" by a group of psychiatrists confronted with a redacted transcript of the Richard Nixon’s Watergate tapes (Schwartz 1982). A system passing Test 4, with the ability to emulate a ‘moderately +’ human, will make for a more human AI, for + within a broad range of usually regarded as unhealthy mental states. We may not use such a system, however, to emulate public elected officials.

**References**

APA. 2013. Diagnostic and Statistical Manual of Mental Disorders (Fifth ed.). American Psychiatric Association.


Williams, J.; Gibbon, M.; First, M.; Spitzer, R.; and et al., M. D. 1992. The structured clinical interview for dsm-iii-r (scid) ii. multi-site test-retest reliability. *Arch Gen Psychiatry* 49:630636.