Combination of Automated Language Analysis with Machine Learning and its Application to Early Diagnosing of Psychotic Disorders

Kamila Gajdka

Jagiellonian University, Institute of Psychology, Ingardena 6, 30-060 Cracow, Poland.

gajdka.kamila@gmail.com

Abstract. Language is often perceived in psychiatry as a window to thoughts processes in humans mind; in clinical assessment disorders of language are usually treated as equivalent of formal thought disorders (FTD). One of the best described in literature connection between language and mental diseases are psychotic disorders. Corcoran et al. [9] identified an automated machine learning speech classifier which can predict psychosis onset in clinical high-risk population with accuracy about 70-80%. Factors which turned out to be the most significant were changes in semantic coherence and syntactic complexity. This method of early diagnosing, apart from its high accuracy, is also more available and have lower costs than other machine-learning-based techniques of diagnosing, which usually requires some neuroimaging data. Furthermore, it is able to give us some deeper look into cognitive processes in psychotic disorders.

Keywords: Automated Language Analysis, Machine Learning, Psychotic Disorders.

1 Introduction

1.1 Schizophrenia and Language – Theoretical Frame

Originally, formal thought disorder (FTD) was concerned to be a core symptom of schizophrenia, usually understood due to classic Bleuler's view, as loosening of associations in thought processes [1, 6]. Later studies found that some kinds of FTD can occur also in other mental disease (for example in mania) and in healthy population as well [1, 2], but there are still some dimensions more which are specifically connected with schizophrenia-related psychosis [2, 3, 12]. especially, negative thought disorders, manifesting in poverty of speech and poverty of content of speech, seems to be strongly connected with this spectrum [3, 12]. Negative thought disorders, in opposition to positive thought disorders, more often occurring in mania, tend to persist in spite of recovery from other symptoms and are connected with poorer outcome in patients who experience them [3]. Additionally, some newer studies have shown that negative thought disorders in mid-childhood could be predictor of later schizophrenia-related psychosis development, while positive thought disorders seems to be more related with affective psychosis[12]. Authors suggests that negative thought disorder could be related to schizotypy, which is thought to be latent personality organization, strongly predicting development of schizophrenia spectrum disorders. [19, 20] There are also some findings which indicates on some language abnormalities, without its association with specific clinical constructs, as in studies previously mentioned, for example different than normal outcome in letter and category fluency tasks [7] or deficits in metaphor comprehension [8]. Specific cognitive and neuronal patterns of language production and processing are observed not only in individuals with diagnosis of psychotic disorders, but also in relatives of patients with this diagnose [18] or individuals who have high scores on schizotypy personality traits [14].

1.2 Machine Learning Approach in Diagnosing Schizophrenia

Machine learning approach is linked with artificial intelligence and it is based on system's ability to learn from its own experience, based on previous analyses of statistical regularities in large set of data [23]. There are some studies, which explore potential of machine-learning-based method of schizophrenia spectrum disorders, most of them based on neuroimaging data [13, 15, 22, 23]. However, there are more studies, which aim to diagnose schizophrenia in individuals, who have already developed some psychotic symptoms, than studies, which goal is to predict psychosis onset in individuals of clinical high-risk [25]. Zarogianni et al. [25] identified method which could predicts psychosis onset in high-risk population with 94% accuracy intra-protocol and 74% cross-protocol [24]. These method required analysis of neuroanatomical data, schizotypal personality traits and some specific neurocognitive features [24, 25]. However, methods which requires to collect some neuroimaging data are usually expensive and not available for every researcher or clinician.

1.3 Automated Methods of Language Analysis

One of the first attempt to identifying automated methods of language analysis and its application in diagnosing FTD were studies of Elvevag et al. [10, 11]. In their first study, they used some Latent Semantic Analysis (LSA) to assess differences in coherence of discourse between groups of patients with diagnosis of schizophrenia and healthy controls. LSA is computational method of text analyzing, considering specific approach to human semantic knowledge acquisition, which assumes that meaning of the word is inferring and learned in accordance to its co-occurrence with other words in text [16, 17]. Elvevag et al. [11] found significant differences between groups and significant correlation between their method and clinical ratings of FTD [11]. In their second study, using the same method, they have observed similar effects in first-degree relatives of patients with schizophrenia [10]. In later study, Bedi et al.[5] combined automated language analysis with Machine Learning to identify system, which could be able to predict later psychosis onset in clinical high-risk youth. They found some speech features, which occurred to predict psychosis onset with accuracy of 100%, although, their study was conducted on very small group [5].

2 Referred Study

Corcoran et al [9] used some machine learning algorithm to identify automated natural language processing method, which would be able to predict psychosis onset in clinical high-risk youth. Process of Machine Learning is based on computers analysis of large amount of data, in this case, large corpus of text, in aim to systems acquisition of vocabulary (semantic) and grammar (syntax). In acquisition of semantic, Corcoran et al. [9] used Latent Semantic Analysis (LSA) and for acquisition of syntax, there was used part-of-speech tagging method, which is able to determine length of sentences and rates usage of different parts of speech [21]. First part of study, included prompt-based dataset protocols from study of Bearden et al [4]. This dataset had been used to train systems algorithm in speech classification and to study intra-protocol method accuracy. In the second part, were used narrative-based dataset protocols from Bedi et al [5]. Machine Learning algorithm was aimed to

classify speech by characteristics of these who developed later psychosis, compared to this who did not. Machine Learning process was circumscribed to eleven speech variables which had differ between CHR+ group and CHR- group in study of Bearden et al [4] and three variables from Bedi et al [5]. Identified characteristics, which discriminate clinical high-risk group, who developed later psychosis, from these who did not, occurred to be decreased semantic coherence, greater variance in that coherence and reduced usage of possessive pronouns. These characteristics had 83% accuracy in predicting psychosis onset intra-protocol (training dataset), a cross-validated 79% (test dataset) accuracy in predicting psychosis onset in the original high-risk cohort (cross-protocol) and 72% accuracy in discriminating speech of recent-onset psychotic patients from healthy individuals [9]. In both studies have been also created some convex hull classifications in which speech data points from non-converters were inside a hull, while those from converters were outside a hull. Similar hull was created for comparison of healthy controls with recent-onset psychotic patients. In this case, data points from patients was largely outside the hull. These findings suggests, that language of pre-psychotic and psychotic individuals is significantly deviant from a constrained hull of language of healthy individuals in aspects of both semantics and syntax [9].

3 Summary

Referred study have shown some potential of using automated methods of predicting psychosis onset in clinical high-risk youth. Combination of automated speech analysis with Machine Learning , in opposition to Machine Learning methods based on neuroimaging data, have some advantage of its availability and lower costs. Additionally, taking some deeper attention to language processing in psychotic disorders and their prodromal phase, could give us a greater insight to cognitive processes underlying their pathology and stronger bases to improving therapeutic methods.

References

1. Andreasen, N.C.: Thought, language and communication disorders. I. Clinical assessment, definition of terms, and evaluation of their reliability. Archives of General Psychiatry 36, 1315-1321 (1979).
2. Andreasen, N.C.: Thought, language and communication disorders. II. Diagnostic significance. Archives of General Psychiatry 36, 1325-1330 (1979).
3. Andreasen, N.C., Grove, W.M.: Thought, language and communication in schizophrenia: Diagnosis and prognosis. Schizophrenia Bulletin 12(3), 348-359 (1986).
4. Bearden, C.E., Nei Wu, K., Caplan, R., Cannon, T.D.: Thought disorder and communication deviance as predictors of outcome in youth at clinical high risk for psychosis. Journal of American Academic Child and Adolescent Psychiatry 50(7), 669-680 (2011).
5. Bedi, G., Carrillo, F., Cecchi, G.A., Slezak, D.F., Sigman, M., Mota, N.B., Ribeiro, S., Javitt, D.C., Copelli, M., Corcoran, C.M.: Automated analysis of speech predicts psychosis onset in high-risk youth. Npj Schizophrenia 1, 15030 (2015). doi:10.1038/npjschz.2015.30
6. Bleuler, E.: Dementia Praecox, or The Group of Schizophrenias, Zinkin, J. (trans). New York, International Universities Press (1950).
7. Bokat, Ch.E., Goldberg, T.E.: Letter and category fluency in schizophrenic patients: a meta-analysis. Schizophrenia Research 64, 73-78 (2003). doi:10.1016/S0920-9964(02)00282-7
8. Chakrabarty, M., Sarkar, S., Chatterjee, A., Ghosal, M., Guha, P., Deogaonkar, M.: Metaphor comprehension deficit in schizophrenia with reference to the hypothesis of abnormal lateralization and right hemisphere dysfunction. Language Sciences 44, 1-14 (2014). http://dx.doi.org/10.1016/j.langsci.2014.01.002
9. Corcoran, C.M., Carrillo, F., Fernandez-Slezak, D., Bedi, G., Kilm, C., Javitt, D.C., Bearden, C.E., Cecchi, C.A.: Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry 17, 67-75 (2018). doi:10.1002/wps.20491
10. Elvevag, B., Foltz, P.W., Rosenstein, M., De Lisi, L.E.: An Automated Method to Analyze Language Use in Patients with Schizophrenia and Their First-Degree Relatives. Journal of Neurolinguistics 23(3), 270-284 (2010). doi:10.1016/j.jneuroling.2009.05.002
11. Elvevag, B., Foltz, P.W., Weinberger, D.R., Goldberg, T.E.: Quantifying Incoherence in Speech: An Automated Methodology and Novel Application to Schizophrenia. Schizophrenia Research 93 (1-3), 304-316 (2007). doi:10.1016/j.schres.2007.03.001
12. Gooding, D.C., Ott S.L., Roberts, S.A., Erlenmeyer-Kimling, L.: Thought disorder in mid-childhood as a predictor of adulthood diagnostic outcome: findings from the New York High-Risk Project. Psychological Medicine 43, 1003-1012 (2013). doi:10.1017/S0033291712001791

13. Greenstein, D., Malley, J.D., Weisinger, B., Clasen, L., Gogtay, N.: Using Multivariate Machine Learning Methods and Structural MRI to Classify Childhood Onset Schizophrenia and Healthy Controls. Frontiers in Psychiatry 3, 53 (2012). doi: 10.3389/fpsyt.2012.00053
14. Kiang, M., Kutas, M.: Association of schizotypy with semantic processing differences: An event-related brain potential study. Schizophrenia Research 77, 329-342 (2005). doi:10.1016/j.schres.2005.03.021
15. Koutsouleris, N., Meisenzahl, E.M., Davatzikos, Ch., Bottlender, R., Frodl, T., Scheuerecker, J., Schmitt, G., Zetzsche, T., Decker, P., Reiser, M., Möller, H.J., Gaser, Ch.: Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. Archives of General Psychiatry 66(7), 700-712 (2009).
16. Landauer, T.K., Dumais, S.T.: A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. Psychological Review 104(2), 211-240 (1997).
17. Landauer, T.K., Foltz, P.W., Laham, D.: An introduction to Latent Semantic Analysis. Discourse Processes 25, 259-284 (1998).
18. Manschreck, T.C., Merrill, A.M., Jabbar, G., Chun, J., DeLisi, L.E. (2012) Frequency of normative word associations in the speech of individuals at familial high-risk for schizophrenia. Schizophrenia Research 140, 99103 (2012). doi:10.1016/j.schres.2012.06.034
19. Meehl, P.E.: Schizotaxia, schizotypy, schizophrenia. American Psychologist 17, 827–838 (1962).
20. Meehl, P.E.: Schizotaxia revisited. Archives of General Psychiatry 46, 935–944 (1989).
21. Santorini, B.: Part-of-speech tagging guidelines for the Penn Treebank Project (3rd Revision). Philadelphia: Department of Computer and Information Science, University of Pennsylvania (1990).
22. Shim, M., Hwang, H.J., Kim, D.W., Lee, S.W., Im, C.H.: Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features. Schizophrenia Research (2016). http://dx.doi.org/10.1016/j.schres.2016.05.007
23. Veronese, E., Castellani, U., Peruzzo, D., Bellani, M., Brambilla, P.: Machine learning approaches: from theory to application in schizophrenia. Computational and Mathematical Methods in Medicine (2013). http://dx.doi.org/10.1155/2013/867924
24. Zarogianni, E., Storkey, A.J., Borgwardt, S., Smieskova, R., Studerus, R., Riecher-Rössler, A., Lawrie, S.M.: Individualized prediction of psychosis in subjects with an at-risk mental state. Schizophrenia Research (2017). http://dx.doi.org/10.1016/j.schres.2017.08.061
25. Zarogianni, E., Storkey, A.J., Johnstone, E.C., Owens, D.G.C., Lawrie, S.M.: Improved individualized prediction of schizophrenia in subjects at familial high risk, based on neuroanatomical data, schizotypal and neurocognitive features. Schizophrenia Research 181, 6-12 (2017). doi:10.1016/j.schres.2007.03.001