Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230161

What Do Autistic People Discuss on Twitter? An Approach Using BERTopic Modelling

Elia GABARRON[a,b 1] and Enrique DORRONZORO[c], Daniel REICHENPFADER[d] and Kerstin DENECKE[d]
[a]*Department of Education, ICT and Learning, Østfold University College, Halden, Norway*
[b]*Norwegian Centre for E-health Research, Tromsø, Norway*
[c]*Universidad de Sevilla, Seville, Spain*
[d]*Bern University of Applied Sciences, Institute for Medical Informatics I4MI, Biel/Bienne, Switzerland*

Abstract. Social media provide easy ways to autistic individuals to communicate and to make their voices heard. The objective of this paper is to identify the main themes that are being discussed by autistic people on Twitter. We collected a sample of tweets containing the hashtag #ActuallyAutistic during the period 10/02/2022 and 14/09/2022. To identify the most discussed topics, BERTopic modelling was applied. We manually grouped the detected topics into 6 major themes using inductive content analysis: 1) General aspects of autism and experiences of autistic individuals; 2) Autism awareness, pride and funding; 3) Interventions, mostly related to Applied Behavior Analysis; 4) Reactions and expressions; 5) Everyday life as an autistic (lifelong condition, work, housing…); and 6) Symbols and characteristics. The majority of tweets were presenting general aspects and experiences as autistic individuals; raising awareness; and about their dissatisfaction with some interventions. The identification of autistic individuals' main discussion themes could help to develop meaningful public health agendas and research involving and addressed to autistic individuals.

Keywords. Autism Spectrum Disorder, Social Media, Twitter, Topic Modelling, BERT

1. Introduction

Autism Spectrum Disorder is characterized for restricted interests, repetitive patterns of behaviors, and impairments in both, socialization and communication [1]. Despite these impairments, many autistic individuals use social media to socialize [2-3], and to make their voices heard [2-4]. Social media are perceived by some autistic people as easier media for communicating and being social than in-person interactions [4].

Not much is known about what autistic individuals discuss on social media. Research identifying and analyzing discussions related to autism on social media exists

[1] Corresponding Author: Elia Gabarron, Department of Education, ICT and Learning, Østfold University College, B R A Veien 4, 1757 Halden (Norway), E-mail: elia.gabarron@hiof.no

[5-8]. However, the analyzed posts by these studies were non-specific and could have been posted by autistic people, or by other individuals interested in that topic. As far as we know, only few studies have specifically focused on messages posted by autistic individuals [9-11]. These studies have found that autistic social media users utilize these channels to make meaning of their own experiences, such as rejecting medicalized narratives or flipping negative narratives into positive stories [9,10]; and to post about autism characteristics and societal barriers [10].

The use of automatic natural language processing methods could help to identify major themes posted by autistic individuals, and therefore increase our understanding of autistic priorities and concerns. The objective of this paper is to identify the main themes that are being discussed by autistic individuals on Twitter by using BERTopic modelling.

2. Method

2.1. Tweets identification and data extraction

Tweets are published within the public domain, and are frequently used in research [5-11]. Twitter provides Application Programing Interface (API), enabling programmatic extraction of the data. Using this API, we collected a sample of tweets including the hashtag #ActuallyAutistic, among others, during the period 10/02/2022 and 14/09/2022. The limitations of the API do only allow to download 900 tweets every 15 minutes. To download a representative quantity of tweets, we developed a script to run a query several times per day. Together with the text of the tweet we collected detailed user metadata (such as user name, profile image, posting time, and follower number). For the purpose of this study, we filtered the collection of tweets to only include those with the hashtag #ActuallyAutistic. This hashtag was chosen for being the most representative one used by social media users as a means of identifying and revealing themselves as autistic.

2.2. Topic modelling and content analysis

We used the BERTopic modelling technique to identify the main discussed themes [12]. Minor data pre-processing included removal of user handles (starting with "@") and hyperlinks, removal of the hashtag "#ActuallyAutistic", and of common English stop words like "the", "a", or "and". Other hashtags were preserved (by removing the "#") and emojis were decoded to corresponding text representations. Furthermore, contractions where expanded ("I'm" -> "I am"). Hence, tweets that only contained user handles and/or hyperlinks vanished and were not considered for analysis. Applying BERTopic modelling comprises following steps: 1) tweets were converted to vector representations; 2) the dimensionality of these vectors was reduced using the UMAP approach [13]; 3) the data was clustered and bag-of-words representations were generated for each cluster; and 4) the topics were generated using a class-based TF-IDF model. As a language embedding model we used the default sentence-transformer model "all-MiniLM-L6-v2" [14]. The code was run using default parameters, except defined as follows: BERTopic (nr_topics=auto, min_topic_size=300), UMAP (n_neigbours=200).

Inductive content analysis was used to group the major categories identified by the BERTopic modelling technique, which included examples of representative tweets, into main themes. Any disagreements were resolved by discussion.

The treatment of personal information for this study was approved by the data protection officer at the University Hospital of North Norway (Ref:02275).

3. Results

A total of 98601 tweets and retweets including the hashtag #ActuallyAutistic were extracted. Authors of these tweets followed an average of 1570,3 other users, and had a mean of 2440,7 followers. After minor pre-processing, a total amount of 97751 tweets was fed into the BERTopic model. The model auto-generated 35 topics related to the Twitter debate including the hashtag #ActuallyAutistic. The number of topics was further auto-reduced by BERTopic to 20. We manually grouped these 20 topics into six major themes using inductive content analysis. Identified major themes were: T1-General aspects and experiences of autistic individuals; T2-Autism awareness, pride and funding; T3-Interventions, mostly related to Applied Behavior Analysis (ABA); T4-Reactions, expressions; T5-Everyday life as an autistic (lifelong condition, work, housing…); and T6-Symbols, characteristics. The remaining tweets were automatically classified by the BERTopic algorithm into two topics including non-English tweets as well as into an "outlier" class, for which topic assignment could not be done. Included topics in these themes, and number of tweets that were classified into these topics are presented in Table 1.

Table 1. Major themes discussed on tweets including the hashtag #ActuallyAutistic

Theme	Sets of relevant words in each theme [BERTtopic grouped sub-topics]	Tweets
T1: General aspects	['autistic', 'autism', 'people', 'like', 'just', 'know', 'neurodivergent', 'adhd', 'autismacceptancemonth' 'autismacceptance']	44214 (45.2%)
T2: Autism awareness, pride, funding	['april', '*oncomingfist*', '*thumbsup*', 'single', '*smilingfacewithsmilingeyes*', '*blueheart*', 'fact', 'autismawarenessmonth', 'year', 'understand'], ['thread', 'april', 'month', 'pure', 'folx', 'know', 'hell', 'probably', 'aware', 'awareness'], ['donate', 'money', 'donating', 'help', 'donations', 'mutualaidrequest', 'autismacceptancemonth', 'gofundme', 'fundraiser', 'support'], ['day', 'choose', 'express', 'pride', 'autisticprideday', 'autistics', 'today', 'autistic', 'aye', 'june'], ['taxi', 'thai', 'nftart', 'nfts', 'autismpride', 'fake', 'nft', 'nftproject', 'nftcommunity', 'art']	3344 (3.4%)
T3: Interventions, mostly ABA	['lgbtq', 'aba', 'applied', 'analysis', 'lovaas', 'behavioral', 'ivar', 'ole', 'fates', 'intertwined'], ['symbols', 'despised', 'refuse', 'eradication', 'dedicated', 'eugenics', 'fund', 'torture', 'company', 'abuse'], ['month', 'awarenessautism', '*foldedhandslightskintone*', 'saynotoautismspeaks', 'clear', 'word', 'sure', 'thank', 'rainbow', 'infinity'], ['cured', 'cure', 'autism', 'disease', 'accommodation', 'abundantly', 'eradicated', 'eugenics', 'acceptance', 'form']	3309 (3.4%)
T4: Reactions, expressions	['description', 'wow', 'perfect', 'problem', 'come', 'answer', 'induction', 'functiooon', 'whomever', 'compelling'], ['funny', 'laugh', 'reasons', 'situation', 'mean', 'actually', 'does', 'person', 'giggle', 'dying'], ['sometimesoften', 'embarrass', 'solid', 'crucial', 'shared', 'details', 'conversation', 'small', 'communication', 'believe']	3302 (3.4%)

T5: Everyday life as an autistic	['lifetime', 'masking', 'trying', 'unmasked', 'standards', 'exhausted', 'holding', 'spent', 'appease', 'concrete'], ['leave', 'functioning', 'checklist', 'lists', 'require', 'jobs', 'house', 'saw', 'low', 'deserve'], ['*hollowredcircle*', '*crossmark*', 'rent', 'help', '164', 'urgent', 'cut', 'fools', 'badly', 'paying']	2081 (2.1%)
T6: Symbols, characteristics	['boys', 'soon', 'created', 'common', 'coming', 'okay', 'thought', 'true', 'puzzle', 'symbol'], ['symbol', 'infinity', 'puzzling', 'outdated', 'gold', 'fixed', 'condition', 'general', 'meaning', 'offensive']	1750 (1.8%)
Non-English tweets	['ich', 'autismus', 'je', 'und', 'et', 'ist', 'pas', 'le', 'die', 'zu'], ['soyautista', 'que', 'el', 'autismo', 'en', 'por', 'autista', 'la', 'autistas', 'para']	1254 (1.3%)
Outliers	Topics that could not be assigned to one of the above-mentioned topics	38497 (39.4%)
Total		**97751**

* *Emojis are presented in italics*

4. Discussion

Many individuals display their conditions on social media, including autistic individuals. The topic analysis approach we used in this paper provides a tool to learn more about aspects of relevance for autistic social media users. Studying topics discussed by autistic people over time can provide details about recent and current pains and needs of those people. In this way, more focused public health agendas and research can be developed.

In our study we have identified that autistic people on Twitter mostly discuss about general aspects and their experiences as autistic individuals, and about raising awareness. Our findings are in agreement with what has been identified in previous research [9-10]. However, dissatisfaction with some current interventions, specifically the use of ABA, was also identified as one of the most main discussed themes in our database (i.e., 'refuse', 'eradication', 'torture', or 'abuse' were some of the most commonly used keywords linked to ABA interventions). Opposition to the use of ABA in education of autistic people has previously been reported as being part of the neurodiversity movement [15-16]. The community disapproval of this type of interventions should be considered by authorities and researchers when developing future intervention programmes.

The literature shows that the neurodiversity movement also claims increased efforts to develop and validate tools to measure autistic prioritized outcomes [16]. The use of automatic natural language processing methods, and specifically BERTopic could help to identify some of these priorities quickly and on time, and by directly listening to users. Both BERTopic and non-negative matrix factorization have been found to be superior techniques for analyzing Twitter data, in comparison to latent Dirichlet allocation and Top2Vec approaches [17]. Advantages of BERTopic include no to little data pre-processing effort, automatic topic reduction and high stability across domains [17]. Moreover, instead of calculating topic distribution, BERTopic assigns each tweet to exactly one topic. However, a potential high number of auto-generated topics and outliers as well as the absence of objective evaluation metrics have also been reported [17].

Our study presents several limitations. We have only used one hashtag as a way to identify messages posted by autistic individuals on Twitter. Findings from this study cannot be generalized to all autistic individuals, neither to all autistic social media users. There are also some limitations to our chosen approach. We decided not to filter non-English tweets from the input data but used an English resulting in the generation of two

"phantom-topics" containing these non-English tweets. As we did not prevent stochastic behaviour when creating the results described in this paper, we are not able to reproduce the exact same topic representations. Several runs of the modelling pipeline however showed similar results across all runs.

5. Conclusions

BERTopic modelling can be a valid approach to identify main themes that are discussed by autistic individuals on Twitter. Autistic people on Twitter seems to mostly discuss about general aspects and experiences as autistic individuals, raising awareness, and about their dissatisfaction with ABA interventions. Future research could further investigate how topic modelling approaches could help to identify autistic individuals' priorities and concerns on time. Since social media has no borders, but public health interventions are developed in specific areas, future research could also investigate how to transfer this knowledge to develop meaningful public health agendas and research that involve and are addressed to autistic individuals in different regions.

References

[1] American Psychiatric Association 2013. Diagnostic and statistical manual of mental disorders. 5th ed.
[2] Ward DM, Dill-Shackleford KE, Mazurek MO. Social Media Use and Happiness in Adults with Autism Spectrum Disorder. Cyberpsychol Behav Soc Netw 2018;21(3):205-209.
[3] Jones S. TikTok is teaching the world about autism – but is it empowering autistic people or pigeonholing them? Theconversation.com 2022. Available at https://theconversation.com/tiktok-is-teaching-the-world-about-autism-but-is-it-empowering-autistic-people-or-pigeonholing-them-192093
[4] Wayman L. Social media offers benefits for autistic community. Autism speaks 2021. Available at: https://www.autismspeaks.org/science-news/social-media-offers-benefits-autistic-community
[5] Shakes P, Cashin A. An analysis of Twitter discourse regarding identifying language for people on the autism spectrum. Issues Ment Health Nurs 2020;51(3):221-228.
[6] Corti L, Zanetti M, et al. Social media analysis of Twitter tweets related to ASD in 2019-2020, with particular attention to COVID-19: topic modelling and sentiment analysis. J Big Data 2022;9:113.
[7] Gabarron E, Dechsling A, Skafle I, et al. Discussions of Asperger Syndrome on Social Media: Content and Sentiment Analysis on Twitter. JMIR Form Res 2022;6(3):e32752.
[8] Skafle I, Gabarron E, Dechsling A, et al. Online Attitudes and Information-Seeking Behaviour on Autism, Asperger Syndrome and Greta Thunberg. Int J Environ Res Public Health 2021;18(9):4981.
[9] Egner J. #ActuallyAutistic: using Twitter to Construct Individual and Collective Identity Narratives. Stud Soc Justice 2022;16(2):349-369.
[10] Zolyomi A, Jones R, Kaftan T. #ActuallyAutistic sense-making on Twitter. Conference on Computers and Accessibility (ASSETS '20), October 26–28,.2020, Virtual Event, Greece.
[11] Hswen Y, et al. Using Twitter to detect psychological characteristics of self-identified persons with Autism Spectrum Disorder: A feasibility study. JMIR Mhealth Uhealth 2019;7(2):e12264.
[12] Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv 2002:2203.05794). Available at: https://doi.org/10.48550/arXiv.2203.05794
[13] McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv 2020:1802.03426. Available at: https://doi.org/10.48550/arXiv.1802.03426
[14] Reimers N, Gurevych I. Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Available at: https://arxiv.org/abs/2004.09813
[15] Løkke JA, Orm S, Dechsling A. Misconceptions about applied behavior analysis in the neurodiversity movement. Norsk Tidsskrift for Atferdsanalyse 2019;46:55-63.
[16] Leadbitter K, Buckle KL, Ellis C, et al. Autistic self-advocacy and the neurodiversity movement: Implications for autism early intervention research and practice. Front Psychol 2021;12:635690.
[17] Egger R, Yu J. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. Front Sociol 2022;7:886498.