A computational linguistic study of personal recovery in bipolar disorder

Abstract

Mental health research can benefit increasingly fruitfully from computational linguistics methods, given the abundant availability of language data in the internet and advances of computational tools. This interdisciplinary project will collect and analyse social media data of individuals diagnosed with bipolar disorder with regard to their recovery experiences. Personal recovery - living a satisfying and contributing life along symptoms of severe mental health issues - so far has only been investigated qualitatively with structured interviews and quantitatively with standardised questionnaires with mainly English-speaking participants in Western countries. Complementary to this evidence, computational linguistic methods allow us to analyse first-person accounts shared online in large quantities, representing unstructured settings and a more heterogeneous, multilingual population, to draw a more complete picture of the aspects and mechanisms of personal recovery in bipolar disorder.

Introduction and background

Recent years have witnessed increased performance in many computational linguistics tasks such as syntactic and semantic parsing BIBREF0 , BIBREF1 , emotion classification BIBREF2 , and sentiment analysis BIBREF3 , BIBREF4 , BIBREF5 , especially concerning the applicability of such tools to noisy online data. Moreover, the field has made substantial progress in developing multilingual models and extending semantic annotation resources to languages beyond English BIBREF6 , BIBREF7 , BIBREF8 , BIBREF9 .

Concurrently, it has been argued for mental health research that it would constitute a `valuable critical step' BIBREF10 to analyse first-hand accounts by individuals with lived experience of severe mental

health issues in blog posts, tweets, and discussion forums. Several severe mental health difficulties, e.g., bipolar disorder (BD) and schizophrenia are considered as chronic and clinical recovery, defined as being relapse and symptom free for a sustained period of time BIBREF11 , is considered difficult to achieve BIBREF12 , BIBREF13 , BIBREF14 . Moreover, clinically recovered individuals often do not regain full social and educational/vocational functioning BIBREF15 , BIBREF16 . Therefore, research originating from initiatives by people with lived experience of mental health issues has been advocating emphasis on the individual's goals in recovery BIBREF17 , BIBREF18 . This movement gave rise to the concept of personal recovery BIBREF19 , BIBREF20 , loosely defined as a `way of living a satisfying, hopeful, and contributing life even with limitations caused by illness' BIBREF18 . The aspects of personal recovery have been conceptualised in various ways BIBREF21 , BIBREF22 , BIBREF23 . According to the frequently used CHIME model BIBREF24 , its main components are Connectedness, Hope and optimism, Identity, Meaning and purpose, and Empowerment. Here, we focus on BD, which is characterised by recurring episodes of depressed and elated (hypomanic or manic) mood BIBREF25 , BIBREF12 . Bipolar spectrum disorders were estimated to affect approximately 2% of the UK population BIBREF13 with rates ranging from 0.1%-4.4% across 11 other European, American and Asian countries BIBREF26 . Moreover, BD is associated with a high risk of suicide BIBREF27 , making its prevention and treatment important tasks for society. BD-specific personal recovery research is motivated by mainly two facts: First, the pole of positive/elevated mood and ongoing mood instability constitute core features of BD and pose special challenges compared to other mental health issues, such as unipolar depression BIBREF25 . Second, unlike for some other severe mental health difficulties, return to normal functioning is achievable given appropriate treatment BIBREF28 , BIBREF16 , BIBREF29 .

A substantial body of qualitative and quantitative research has shown the importance of personal recovery for individuals diagnosed with BD BIBREF22 , BIBREF25 , BIBREF30 , BIBREF31 , BIBREF23 . Qualitative evidence mainly comes from (semi-)structured interviews and focus groups and has been criticised for small numbers of participants BIBREF10 , lacking complementary quantitative evidence from

larger samples BIBREF32 . Some quantitative evidence stems from the standardised bipolar recovery questionnaire BIBREF30 and a randomised control trial for recovery-focused cognitive-behavioural therapy BIBREF31 . Critically, previous research has taken place only in structured settings. What is more, the recovery concept emerged from research primarily conducted in English-speaking countries, mainly involving researchers and participants of Western ethnicity. This might have led to a lack of non-Western notions of wellbeing in the concept, such as those found in indigenous peoples BIBREF32 , limiting its the applicability to a general population. Indeed, the variation in BD prevalence rates from 0.1% in India to 4.4% in the US is striking. It has been shown that culture is an important factor in the diagnosis of BD BIBREF33 , as well as on the causes attributed to mental health difficulties in general and treatments considered appropriate BIBREF34 , BIBREF35 . While approaches to mental health classification from texts have long ignored the cultural dimension BIBREF36 , first studies show that online language of individuals affected by depression or related mental health difficulties differs significantly across cultures BIBREF37 , BIBREF36 .

Hence, it seems timely to take into account the wealth of accounts of mental health difficulties and recovery stories from individuals of diverse ethnic and cultural backgrounds that are available in a multitude of languages on the internet. Corpus and computational linguistic methods are explicitly designed for processing large amounts of linguistic data BIBREF38 , BIBREF39 , BIBREF40 , BIBREF41 , and as discussed above, recent advances have made it feasible to apply them to noisy user-generated texts from diverse domains, including mental health BIBREF42 , BIBREF43 . Computer-aided analysis of public social media data enables us to address several shortcomings in the scientific underpinning of personal recovery in BD by overcoming the small sample sizes of lab-collected data and including accounts from a more heterogeneous population.

In sum, our research questions are as follows: (1) How is personal recovery discussed online by individuals meeting criteria for BD? (2) What new insights do we get about personal recovery and factors

that facilitate or hinder it? We will investigate these questions in two parts, looking at English-language data by westerners and at multilingual data by individuals of diverse ethnicities.

Data

Previous work in computational linguistics and clinical psychology has tended to focus on the detection of mental health issues as classification tasks BIBREF44 . Datasets have been collected for various conditions including BD using publicly available social-media data from Twitter BIBREF45 and Reddit BIBREF46 , BIBREF47 . Unfortunately, the Twitter dataset is unavailable for further research. In both Reddit datasets, mental health-related content was deliberately removed. This allows the training of classifiers that try to predict the mental health of authors from excerpts that do not explicitly address mental health, yet it renders the data useless for analyses on how mental health is talked about online. Due to this lack of appropriate existing publicly accessible datasets, we will create such resources and make them available to subsequent researchers.

We plan to collect data relevant for BD in general as well as for personal recovery in BD from three sources varying in their available amount versus depth of the accounts we expect to find: 1) Twitter, 2) Reddit (focusing on mental health-related content unlike previous work), 3) blogs authored by affected individuals. Twitter and Reddit users with a BD diagnosis will be identified automatically via self-reported diagnosis statements, such as `I was diagnosed with BD-I last week'. To do so, we will extend on the diagnosis patterns and terms for BD provided by BIBREF47 . Implicit consent is assumed from users on these platforms to use their public tweets and posts. SECREF3 Relevant blogs will be manually identified, and their authors will be contacted to obtain informed consent for using their texts.

Since language and culture are important factors in our research questions, we need information on the language of the texts and the country of residence of their authors, which is not provided in a structured

format in the three data sources. For language identification, Twitter employs an automatic tool BIBREF48 , which can be used to filter tweets according to 60 language codes, and there are free, fairly accurate tools such as the Google Compact Language Detector, which can be applied to Reddit and blog posts. The location of Twitter users can be automatically inferred from their tweets BIBREF49 or the (albeit noisy) location field in their user profiles BIBREF50 . Only one attempt to classify the location of Reddit users has been published so far BIBREF51 showing meagre results, indicating that the development of robust location classification approaches on this platform would constitute a valuable contribution. Some companies collect mental health-related online data and make them available to researchers subject to approval of their internal review boards, e.g., OurDataHelps by Qntfy or the peer-support forum provider 7 Cups. Unlike `raw' social media data, these datasets have richer user-provided metadata and explicit consent for research usage. On the other hand, less data is available, the process to obtain access might be tedious within the short timeline of a PhD project and it might be impossible to share the used portions of the data with other researchers. Therefore, we will follow up the possibilities of obtaining access to these datasets, but in parallel also collect our own datasets to avoid dependence on external data providers.

Methodology and Resources

As explained in the introduction, the overarching aim of this project is to investigate in how far information conveyed in social media posts can complement more traditional research methods in clinical psychology to get insights into the recovery experience of individuals with a BD diagnosis. Therefore, we will first conduct a systematic literature review of qualitative evidence to establish a solid base of what is already known about personal recovery experiences in BD for the subsequent social media studies.

Our research questions, which regard the experiences of different populations, lend themselves to several subprojects. First, we will collect and analyse English-language data from westerners. Then, we

will address ethnically diverse English-speaking populations and finally multilingual accounts. This has the advantage that we can build data processing and methodological workflows along an increase in complexity of the data collection and analysis throughout the project.

In each project phase, we will employ a mixed-methods approach to combine the advantages of quantitative and qualitative methods BIBREF52 , BIBREF53 , which is established in mental health research BIBREF54 , BIBREF55 , BIBREF56 , BIBREF57 and specifically recommended to investigate personal recovery BIBREF58 . Quantitative methods are suitable to study observable behaviour such as language and yield more generalisable results by taking into account large samples. However, they fall short of capturing the subjective, idiosyncratic meaning of socially constructed reality, which is important when studying individuals' recovery experience BIBREF59 , BIBREF22 , BIBREF23 , BIBREF60 . Therefore, we will apply an explanatory sequential research design BIBREF53 , starting with statistical analysis of the full dataset followed by a manual investigation of fewer examples, similar to `distant reading' BIBREF61 in digital humanities.

Since previous research mainly employed (semi-)structured interviews and we do not expect to necessarily find the same aspects emphasised in unstructured settings, even less so when looking at a more diverse and non-English speaking population, we will not derive hypotheses from existing recovery models for testing on the online data. Instead, we will start off with exploratory quantitative research using comparative analysis tools such as Wmatrix BIBREF62 to uncover important linguistic features, e.g., on keywords and key concepts that occur with unexpected frequency in our collected datasets relative to reference corpora. The underlying assumption is that keywords and key concepts are indicative of certain aspects of personal recovery, such as those specified in the CHIME model BIBREF24 , other previous research BIBREF22 , BIBREF23 , BIBREF60 , or novel ones. Comparing online sources with transcripts of structured interviews or subcorpora originating from different cultural backgrounds might uncover aspects that were not prominently represented in the accounts studied in prior research.

A specific challenge will be to narrow down the data to parts relevant for personal recovery, since there is no control over the discussed topics compared to structured interviews. To investigate how individuals discuss personal recovery online and what (potentially unrecorded) aspects they associate with it, without a priori narrowing down the search-space to specific known keywords seems like a chicken-and-egg problem. We propose to address this challenge by an iterative approach similar to the one taken in a corpus linguistic study of cancer metaphors BIBREF63 . Drawing on results from previous qualitative research BIBREF24 , BIBREF23 , we will compile an initial dictionary of recovery-related terms. Next, we will examine a small portion of the dataset manually, which will be partly randomly sampled and partly selected to contain recovery-related terms. Based on this, we will be able to expand the dictionary and additionally automatically annotate semantic concepts of the identified relevant text passages using a semantic tagging approach such as the UCREL Semantic Analysis System (USAS) BIBREF64 . Crucially for the multilingual aspect of the project, USAS can tag semantic categories in eight languages BIBREF8 . Then, semantic tagging will be applied to the full corpus to retrieve all text passages mentioning relevant concepts. Furthermore, distributional semantics methods BIBREF65 , BIBREF66 can be used to find terms that frequently co-occur with words from our keyword dictionary. Occurrences of the identified keywords or concepts can be quantified in the full corpus to identify the importance of the related personal recovery aspects.

Linguistic Inquiry and Word Count (LIWC) BIBREF67 is a frequently used tool in social-science text analysis to analyse emotional and cognitive components of texts and derive features for classification models BIBREF47 , BIBREF46 , BIBREF68 , BIBREF69 . LIWC counts target words organised in a manually constructed hierarchical dictionary without contextual disambiguation in the texts under analysis and has been psychometrically validated and developed for English exclusively. While translations for several languages exist, e.g., Dutch BIBREF9 , and it is questionable to what extent LIWC concepts can be transferred to other languages and cultures by mere translation. We therefore aim to apply and develop methods that require less manual labour and are applicable to many languages and cultures.

One option constitute unsupervised methods, such as topic modelling, which has been applied to explore cultural differences in mental-health related online data already BIBREF37 , BIBREF36 . The Differential Language Analysis ToolKit (DLATK) BIBREF70 facilitates social-scientific language analyses, including tools for preprocessing, such as emoticon-aware tokenisers, filtering according to meta data, and analysis, e.g. via robust topic modelling methods.

Furthermore, emotion and sentiment analysis constitute useful tools to investigate the emotions involved in talking about recovery and identify factors that facilitate or hinder it. There are many annotated datasets to train supervised classifiers BIBREF71 , BIBREF3 for these actively researched NLP tasks. Machine learning methods were found to usually outperform rule-based approaches based on look-ups in dictionaries such as LIWC. Again, most annotated resources are English, but state of the art approaches based on multilingual embeddings allow transferring models between languages BIBREF4 .

Ethical considerations

Ethical considerations are established as essential part in planning mental health research and most research projects undergo approval by an ethics committee. On the contrary, the computational linguistics community has started only recently to consider ethical questions BIBREF72 , BIBREF73 . Likely, this is because computational linguistics was traditionally concerned with publicly available, impersonal texts such as newspapers or texts published with some temporal distance, which left a distance between the text and author. Conversely, recent social media research often deals with highly personal information of living individuals, who can be directly affected by the outcomes BIBREF72 .

BIBREF72 discuss issues that can arise when constructing datasets from social media and conducting analyses or developing predictive models based on these data, which we review here in relation to our project: Demographic bias in sampling the data can lead to exclusion of minority groups, resulting in

overgeneralisation of models based on these data. As discussed in the introduction, personal recovery research suffers from a bias towards English-speaking Western individuals of white ethnicity. By studying multilingual accounts of ethnically diverse populations we explicitly address the demographic bias of previous research. Topic overexposure is tricky to address, where certain groups are perceived as abnormal when research repeatedly finds that their language is different or more difficult to process. Unlike previous research BIBREF45 , BIBREF47 , BIBREF46 our goal is not to reveal particularities in the language of individuals affected by mental health problems. Instead, we will compare accounts of individuals with BD from different settings (structured interviews versus informal online discourse) and of different backgrounds. While the latter bears the risk to overexpose certain minority groups, we will pay special attention to this in the dissemination of our results.

Lastly, most research, even when conducted with the best intentions, suffers from the dual-use problem BIBREF74 , in that it can be misused or have consequences that affect people's life negatively. For this reason, we refrain from publishing mental health classification methods, which could be used, for example, by health insurance companies for the risk assessment of applicants based on their social media profiles.

If and how informed consent needs to be obtained for research on social media data is a debated issue BIBREF75 , BIBREF76 , BIBREF77 , mainly because it is not straightforward to determine if posts are made in a public or private context. From a legal point of view, the privacy policies of Twitter and Reddit, explicitly allow analysis of the user contents by third party, but it is unclear to what extent users are aware of this when posting to these platforms BIBREF78 . However, in practice it is often infeasible to seek retrospective consent from hundreds or thousands of social media users. According to current ethical guidelines for social media research BIBREF79 , BIBREF80 and practice in comparable research projects BIBREF81 , BIBREF78 , it is regarded as acceptable to waive explicit consent if the anonymity of the users is preserved. Therefore, we will not ask the account holders of Twitter and Reddit posts included in

our datasets for their consent.

BIBREF79 formulate guidelines for ethical social media health research that pertain especially to data collection and sharing. In line with these, we will only share anonymised and paraphrased excerpts from the texts, as it is often possible to recover a user name via a web search for the verbatim text of a post. However, we will make the original texts available as datasets to subsequent research under a data usage agreement. Since the (automatic) annotation of demographic variables in parts of our dataset constitutes especially sensitive information on minority status in conjunction with mental health, we will only share these annotations with researchers that demonstrate a genuine need for them, i.e. to verify our results or to investigate certain research questions.

Another important question is in which situations of encountering content indicative of a risk of self-harm or harm to others it would be appropriate or even required by duty of care for the research team to pass on information to authorities. Surprisingly, we could only find two mentions of this issue in social media research BIBREF81 , BIBREF82 . Acknowledging that suicidal ideation fluctuates BIBREF83 , we accord with the ethical review board's requirement in BIBREF81 to only analyse content posted at least three months ago. If the research team, which includes clinical psychologists, still perceives users at risk we will make use of the reporting facilities of Twitter and Reddit.

As a central component we consider the involvement of individuals with lived experience in our project, an aspect which is missing in the discussion of ethical social media health research so far. The proposal has been presented to an advisory board of individuals with a BD diagnosis and was received positively. The advisory board will be consulted at several stages of the project to inform the research design, analysis, and publication of results. We believe that board members can help to address several of the raised ethical problems, e.g., shaping the research questions to avoid feeding into existing biases or overexposing certain groups and highlighting potentially harmful interpretations and uses of our results.

Impact and conclusion

The importance of the recovery concept in the design of mental health services has recently been prominently reinforced, suggesting ‘recovery-oriented social enterprises as key component of the integrated service’ BIBREF20 . We think that a recovery approach as leading principle for national or global health service strategies, should be informed by voices of individuals as diverse as those it is supposed to serve. Therefore, we expect the proposed investigations of views on recovery by previously under-researched ethnic, language, and cultural groups to yield valuable insights on the appropriateness of the recovery approach for a wider population. The datasets collected in this project can serve as useful resources for future research. More generally, our social-media data-driven approach could be applied to investigate other areas of mental health if it proves successful in leading to relevant new insights.

Finally, this project is an interdisciplinary endeavour, combining clinical psychology, input from individuals with lived experience of BD, and computational linguistics. While this comes with the challenges of cross-disciplinary research, it has the potential to apply and develop state-of-the-art NLP methods in a way that is psychologically and ethically sound as well as informed and approved by affected people to increase our knowledge of severe mental illnesses such as BD.

Acknowledgments

I would like to thank my supervisors Steven Jones, Fiona Lobban, and Paul Rayson for their guidance in this project. My heartfelt thanks go also to Chris Lodge, service user researcher at the Spectrum Centre, and the members of the advisory panel he coordinates that offer feedback on this project based on their lived experience of BD. Further, I would like to thank Masoud Rouhizadeh for his helpful comments during pre-submission mentoring and the anonymous reviewers. This project is funded by the Faculty of Health and Medicine at Lancaster University as part of a doctoral scholarship.