NLG-Based Moderator Response Generator to Support Mental Health

M. Sazzad Husain
Faculty of Engineering & IT
The University of Sydney
New South Wales, Australia
sazzad.hussain@sydney.edu.au

Juchen Li
Faculty of Engineering & IT
The University of Sydney
New South Wales, Australia
juli4748@uni.sydney.edu.au

Rafael A. Calvo
Faculty of Engineering & IT
The University of Sydney
New South Wales, Australia
Rafael.calvo@sydney.edu.au

Laura Ospina-Pinillos
Faculty of Medicine
The University of Sydney
New South Wales, Australia
lauraospin@gmail.com

Louise Ellis
Faculty of Medicine
The University of Sydney
New South Wales, Australia
louise.ellis@sydney.edu.au

Tracey Davenport & Ian Hickie
Faculty of Medicine
The University of Sydney
New South Wales, Australia
tracey.davenport@sydney.edu.au
ian.hickie@sydney.edu.au

Abstract
The global need to effectively address mental health problems and wellbeing is well recognised. Today, online systems are increasingly being viewed as an effective solution for their ability to reach broad populations. As online support groups become popular the workload for human moderators increases. Maintaining quality feedback becomes increasingly challenging as the community grows. Tools that can automatically detect mental health problems from social media posts and then generate smart feedback can greatly reduce human overload. In this paper, we present a system for the automation of interventions using Natural Language Generation (NLG) techniques. In particular, we focus on 'depression' and 'anxiety' related interventions. Psychologists evaluated the quality of the systems' interventions and results were compared against human (i.e. moderator) interventions. Results indicate our intervention system still has a long way to go, but is a step in the right direction as a tool to assist human moderators with their service.

Author Keywords
Mental health, wellbeing, intervention, NLG, HCI

Introduction

There is international recognition of the premature death and disability costs attributable to common mental health disorders [1]. Almost half of all Australians aged 16 to 85 years experience a mental disorder at some point in their life [2]; and, 14% of young Australians have a mental health problem of which just 25% of receive professional help [3]. These statistics are common for most countries.

Despite its seriousness, mental heath has long been a neglected field, particularly for vulnerable and rural populations where access to quality care is limited. Generally speaking, this field struggles with lack of appropriate recognition and diagnosis by health professionals, inappropriate help-seeking in the general population [4-6], strong stigmatisation of mental illness, high costs, and for some communities multiculturalism and language barriers [4].

In order to address these difficulties, computerised and Internet-based interventions have become popular [5, 6]. The simplest online intervention is static information. Under this modality people can find material about their mental health problem by searching keywords and browsing them on the World Wide Web. Social networks also allow people to interact with others experiencing similar problems and they can obtain peer-to-peer support [7]. In addition, people who participate with online mental health groups (e.g., ReachOut.com) can obtain assistance from professional human moderators. With the use of online support groups gaining in popularity, the workload for moderators may increase faster than the resources, directly affecting the quality of the support obtained and potentially making the communities unsustainable.

Systems that apply automatic techniques present a novel and cost-effective solution to detect mental health problems from social media posts and eventually generate quality interventions that can help users and also assist human moderators and peers [7]. As part of generating quality interventions automatically, Natural Language Generation (NLG) techniques can be very suitable. In contrast to the simple, static, rigid, repetitive and impersonal template based feedback; NLG can produce dynamic human-like, individualised sentence structures suitable to various contexts. These high quality interventions can be generated by combining psychological approaches such as cognitive behavioural therapy (CBT) [8] with the information obtained from the posts (e.g., mental health problem) and the timestamps (e.g., posts received, interventions sent).

The aim of this project is to develop an NLG-based system that will create appropriate automatic interventions using input from mental health professionals. The interventions can then be administered by human moderators and can be delivered directly to individuals though social media or online forums. This paper details the results from a pilot study, which directly compares computer-generated interventions to human-generated interventions. Sample of posts ($n=25$) were randomly selected from various mental health forums to generate the interventions. We then asked professional moderators from ReachOut.com to provide human interventions for the same posts. Finally, three academic and clinical psychologists rated the interventions using six quality measures.

Background and NLG architecture for Mental Health Interventions

NLG is a subfield of artificial intelligence and computational linguistics, which primarily focuses on producing human-understandable texts from nonlinguistic data [9]. The field of NLG started to become diverse in the late 1990s and several NLG systems were developed with a growing number of real-word applications [9-12]. However, the development of NLG systems for personalised interventions in the mental health domain is still very fresh. There are different possible architectures for NLG systems, but the one proposed by Reiter and Dale [9] is broadly compatible with most applications.

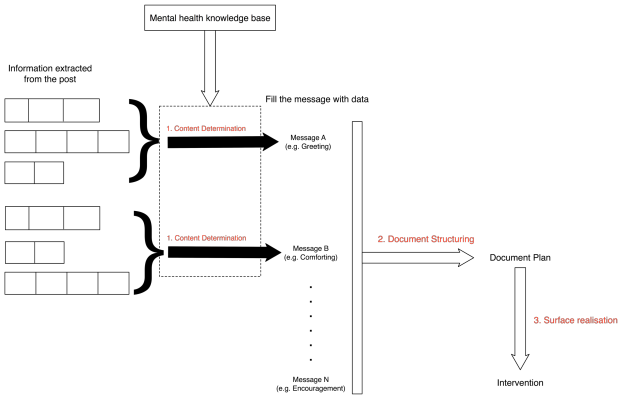

Figure 1. Overview of the NLG Architecture for Mental Health Interventions

Our Moderator Assistant [7] system will be responsible for extracting mental health categories (e.g., depression, anxiety) from social media posts and provide relevant resources (e.g., web links). The input for the NLG system would be the mental health categories, the resources and potentially other information (e.g., age, medication, social behaviour, etc.) extracted from the posts. Figure 1 gives the overview of the NLG architecture described next.

The first step in the architecture is *Content Determination*. The extracted mental health categories are mapped into corresponding intervention information using the knowledge base to construct *Messages*. Each *Message* represented a chunk of data that can be grouped together to express a specific meaning. The second step is *Document Structuring*, where schema and heuristic algorithm are used to combine messages into a *Document Plan*. This is then passed to the surface *Realiser*, where the *Document Plan* is converted into real text from the abstract representations. The system will then produce the intended interventions.

Defining Message

In one approach, *Message* is defined by grouping sets of information that need to be described together. Based on interventions templates provided by psychologists and samples feedback (i.e. comments) found in moderated health support groups, we have identified four types of *Messages* that need to be described together as of the mental health intervention system: (a) *Greeting Message*, (b) *Comforting Message*, (c) *Suggestion Message*, and (d) *Encouragement Message*. The intervention templates and the sample feedback provide well-structured complete sentences that are used as contents for *Messages*.

Quality Measure Questions

Q1: The intervention is gramatically correct?

Q2: The intervention is clear and unambiguous?

Q3: The intervention is appropriate?

Q4: The intervention provides useful resources based on the mental health problem?

Q5: The intervention expresses compassion and warmth?

Q6: The intervention is likely to provide encouragement towards enhancing mental health and wellbeing?

Feedback Generation

The *Content Determination* stage constructs *Messages* based on the input mental health categories. The greeting messages are generated based on the current system time. Then the messages for comforting, suggestion, and encouragement are generated based on the category of the mental health problem. Messages are then used as input for *Document Structuring*. Finally, the *Realiser* constructs the intervention by traversing the *Document Plan* tree [9].

Pilot Evaluation

This section presents the pilot evaluation for the intervention system in the context of depression and anxiety. The project as part of the Moderator Assistant [7] is approved by The University of Sydney Human Research Ethics Committee.

We have collected sample posts from online peer-to-peer and moderated health support groups in *Livejournal*, *Facebook*, *ReachOut.com*. For each post, the author's name (i.e. username) and identifying information was removed. For the evaluation, 25 posts were selected at random from the larger corpora we have collected. Academic and clinical psychologists/psychiatrists classified these posts under the categories of depression, anxiety or both.

The 25 posts were used to generate the interventions using the architecture described in section 2. A professional moderator from ReachOut.com provided the human interventions for the selected posts. The posts were presented to the moderator with the respective categories using a simple web-based application and interventions were collected as comments. The total set of 50 interventions for the 25 posts were randomised and then presented for the rating procedure.

The three psychologists/psychiatrists rated the posts. We identify them as R1, R2, and R3 in this paper for presenting the results. Using a Likert scale (Strongly Disagree, Disagree, Some of the time, Agree, Strongly Agree) the six questions were asked to measure the quality.

For each intervention, raters were also asked in a web application to state whether a computer generated them or a human. Each item included the original post, the intervention and the quality measure assessment tool. As a part of the study, raters knew that the posts were originally annotated under depression or anxiety. However the actual category for each post was never revealed. Additionally, information about the author (human or system) of the interventions was concealed.

Results and Discussion

Firstly, we report the quality of the system interventions using the rating scores. The Likert scale was converted to 1-5 rating scales for estimates. Figure 2 gives the average rating scores for *anxiety*, *depression* and *both*.

Q1 received the highest average rating (slightly above 4), followed by Q2 (above 3). All other questions received average ratings below 3 (expect *depression* for Q4). This indicates that the interventions were grammatically correct and clear to read. *Anxiety* and *depression* interventions were also capable of delivering useful resources (Q4). Standard deviations were high for some questions in some categories, indicating variations in rating scores.

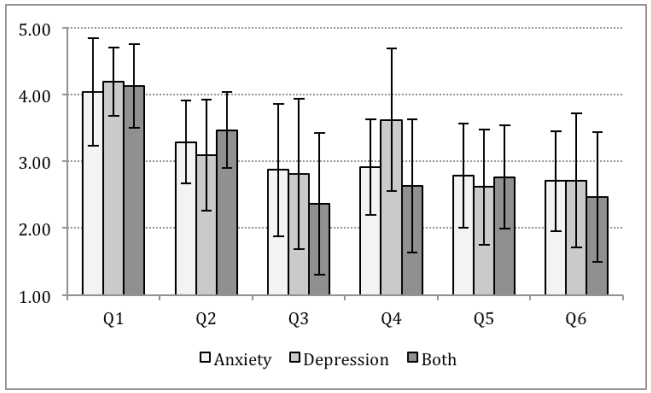

Figure 2. Average rating scores for system interventions.

Even though the average performance for other questions were comparatively lower, majority of the interventions (except *both* for Q3) received ratings above 2. For Q1 (all categories), majority of the interventions scored above 3.

Secondly, we compare the performance of the system interventions to human ones. Tables 1-3 give the average and standard deviation scores for the three raters over all interventions. The scores received by rater R2 indicate very similar quality for both human and system. Raters R1 and R3 scored similarly for human and system in Q1, followed by Q2, however a similar trend as Figure 2 follows for the other questions. The standard deviations indicate similar variations in ratings for both human and system interventions.

It is useful to report the maximum rating scores. Human interventions have received the maximum score of 5 (by R1 and R3) for all the questions. The maximum scores for system interventions are mostly 4 with some 5 over all questions by all raters. This indicates that the system is capable of generating good quality interventions for all questions, but the performance varies over many samples.

R1	*Q1*	*Q2*	*Q3*	*Q4*	*Q5*	*Q6*
Hum	3.84	4.68	4.72	4.24	4.24	4.20
	(1.03)	*(0.56)*	*(0.68)*	*(0.72)*	*(0.52)*	*(0.82)*
Sys	3.96	3.32	2.04	2.36	2.76	1.84
	(0.45)	*(0.80)*	*(1.02)*	*(1.11)*	*(0.78)*	*(0.85)*

Table 1. Comparing interventions ratings (Rater 1)

R2	*Q1*	*Q2*	*Q3*	*Q4*	*Q5*	*Q6*
Hum	3.32	3.32	3.40	3.08	3.44	3.20
	(0.75)	*(0.75)*	*(0.50)*	*(0.64)*	*(0.71)*	*(0.50)*
Sys	3.68	3.52	3.56	3.60	3.28	3.24
	(0.56)	*(0.51)*	*(0.51)*	*(0.65)*	*(0.54)*	*(0.52)*

Table 2. Comparing interventions ratings (Rater 2)

R3	*Q1*	*Q2*	*Q3*	*Q4*	*Q5*	*Q6*
Hum	4.64	4.52	4.32	4.24	4.36	3.88
	(0.57)	*(0.65)*	*(0.90)*	*(0.83)*	*(0.86)*	*(0.83)*
Sys	4.72	3.08	2.36	3.04	2.16	2.76
	(0.46)	*(0.64)*	*(0.95)*	*(0.84)*	*(0.62)*	*(0.72)*

Table 3. Comparing interventions ratings (Rater 3)

In regards to the question whether the interventions were auto generated by a computer, R2 assumed that only 36% of the system interventions were auto generated. For this rater the system interventions appeared to be mostly natural and human-like. However, both R1 and R2 assumed that majority (96%) of the interventions were auto generated. In contrary,

some of the human interventions (10% in total) were also perceived to be auto generated.

Conclusion and Future Work

This paper proposes an NLG-based system for generating natural language interventions for supporting mental health. Despite the variations in rating scores, the performance of the systems interventions for depression and anxiety were satisfactory as part of the early development and pilot evaluation. The results were good indicative of the capability of the system for generating natural language interventions, but needs improvement in sustaining its quality. Mental health interventions should always be moderated, so even with this level of performance the system can greatly reduce human workload. Inclusion of other mental health categories and extraction of more information from posts to improve the NLG knowledge base will be part of future work. The evaluation will be conducted more extensively in the future as well.

Acknowledgements

This project is supported by the Young and Well Cooperative Research Centre, which is established under the Australian Government's Cooperative Research Centres Program. We would like to thank the moderator from ReachOut.com for providing the human interventions and the BMRI for rating the interventions.

References

[1] *The world health report 2001 - Mental Health: New Understanding, New Hope*, in *World Health Organization*. 2001.

[2] *National survey of mental health and wellbeing: Summary of results*, in *Australian Bureau of Statistics*. 2007, Australian Bureau of Statistics Canberra.

[3] Sawyer, M.G., et al., *The mental health of young people in Australia: key findings from the child and adolescent component of the national survey of mental health and well-being.* Australian and New Zealand Journal of Psychiatry, 2001. **35**(6): p. 806-814.

[4] Clarke, G. and B.J. Yarborough, *Evaluating the promise of health IT to enhance/expand the reach of mental health services.* General Hospital Psychiatry, 2013. **35**(4): p. 339-344.

[5] Burns, J.M., et al., *The internet as a setting for mental health service utilisation by young people.* Medical Journal of Australia, 2010. **192**(11):p. S22-S26

[6] Strecher, V., *Internet methods for delivering behavioral and health-related interventions (eHealth).* Annu. Rev. Clin. Psychol., 2007. **3**: p. 53-76.

[7] Liu, M., et al., *Moderator Assistant: helping those who help via online mental health support groups*, in *Joint Workshop on Smart Health and Social Therapies, OzChi2013*. 2013: Adelaide, Australia.

[8] Van Bilsen, H., *Cognitive behaviour therapy in the real world: Back to basics*. 2013: Karnac Books.

[9] Reiter, E., R. Dale, and Z. Feng, *Building natural language generation systems*. Vol. 33. 2000: MIT.

[10] Reiter, E. *Natural Language Generation in STOP*. 1999 [cited 2014; Available from: http://inf.abdn.ac.uk/research/stop/stop-nlg.htm.

[11] DiMarco, C., et al. *The development of a natural language generation system for personalized e-health information*. in *12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems*. 2007. IOS Press.

[12] Varges, S., et al. *SemScribe: Natural Language Generation for medical reports*. in *Eight International Conference on Language Resources and Evaluation (LREC)*. 2012. Istanbul, Turkey.