Using Topic Modelling to Personalise a Digital Self-compassion Training

Laura M. van der Lubbe(✉), Nina Groot, and Charlotte Gerritsen

Computer Science, Vrije Universiteit Amsterdam, De Boelelaan 1111, 1081 HV Amsterdam, The Netherlands
l.m.vander.lubbe@vu.nl

Abstract. Young adults that struggle with mental health issues experience barriers to seek help. With our online self-compassion training we try to overcome some of these barriers. To improve our training, we can personalise exercises based on topic modelling. Data from a pilot study is used to analyse and evaluate the algorithm. Overall, the algorithm has an accuracy of 54.1% for predicting the right topic. This accuracy increases to 80.4% when considering an empty prediction to be correct as well. Although this research also shows that our data makes the task of topic modelling difficult, it does prove to be a possibility to personalise the designed training.

Keywords: Self-compassion · Mental health · Personalization · Topic modelling

1 Introduction

Young adults can struggle with mental health issues due to various reasons. Mental health disorders lead to a poor quality of life and have a high contribution to the global burden of disease [11]. Although young adults are struggling with their mental health, they perceive several barriers to seek help [5]. Among these barriers are the lack of accessibility, preference for self-help instead of external help, and a lack of knowledge of what services exist.

To overcome the barriers around accessibility and self-help, an online intervention is a promising tool. It has the convenience that it is a private and flexible way of training yourself. We designed an online self-compassion training with gamification elements for young adults [7]. Self-compassion means that you are kind to yourself in difficult times, you perceive your experiences as part of the larger human experience, and that you have a mindful attitude towards difficult emotions. It has been associated with positive outcomes, such as an association with greater happiness, optimism, and positive affect [9]. Often, training courses for self-compassion are in person, in which the trainer can have interaction with the participants [3]. However, online alternatives also exist [4,10].

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2022
Published by Springer Nature Switzerland AG 2022. All Rights Reserved
H. Lewy and R. Barkan (Eds.): PH 2021, LNICST 431, pp. 522–532, 2022.
https://doi.org/10.1007/978-3-030-99194-4_32

The online self-compassion training we designed is self-guided, meaning there is no professional supervision or guidance during the training. While in in-person self-compassion training such personal guidance plays an important role, we have to create an automated alternative that still provides some personalization e.g. within one of the exercises [7]. During the pilot study, that we performed to assess the user-experience of the website, we discussed the option of personalising the training content with the participants [8]. By adapting the exercises of the training to topics that participants discuss in different components of the training, it becomes possible to better suit the exercises to the needs of individual users. Therefore we have developed an algorithm for topic modelling that can be used in the improved version of our online self-compassion training.

This paper first discusses background literature on topic modelling. Next, the existing online self-compassion training is introduced and the method of the data analysis and algorithm implementation are described. Followed by the results of this analysis and the evaluation of the algorithm. Finally, we will draw conclusions on how this algorithm can be used to personalise the content of our training website, and what lessons can be learned from a technical perspective.

2 Topic Modelling

Topic modelling means finding themes in unstructured documents [2]. Different approaches and algorithms for this task exist. In [6], different ways of classifying these algorithms are described. First, the classification based on used strategy: probabilistic or non-probabilistic (or algebraic models). Non-probabilistic models use a Bag-of-Words (BoW) approach. In this approach the corpus gets converted into a term document matrix and the order of terms is neglected. The probabilistic model improves such non-probabilistic models by adding the probability sense using generative model approaches. The next classification that can be made is that of supervised and unsupervised approaches. The main difference between these two approaches is the existence of labels in the training data set [1]. Supervised modelling works with predetermined output attributes (labels). The models attempt to predict and classify the predetermined attribute, and their accuracies (alongside other performance measures) is dependent on the number of correctly classified attributes. Unsupervised modelling, on the other hand, focuses on clustering without the use of target attributes. Lastly, one can distinct whether algorithms use the sequence of words during topic modelling or use the BoW approach that does not consider this [6].

3 Method

Currently, the online self-compassion training consists of three exercises, a journal consisting of three components, and a profile page [7]. Gamification is added in the form of a story that the user progresses through. This story is about a journey that you are making: a metaphor for your self-development through learning about self-compassion. The story is a way to deliver theory to the user in a recognisable context. Moreover, to progress on your journey you have to

earn kilometres, which you earn by engaging in exercises and the different journals. For each finished component you progress a certain number of kilometres, and when completing all components on a day you earn a bonus. There is a maximum progression per day, as more interaction is not considered beneficial anymore.

Initially, the story and exercises use situations written around the topics of social media, body image issues, social anxiety and troubles with friends, with as main goal reducing body image related issues of young adults. With assigning topics to previous text written by users, the choice for those situations can be personalised to the individual user. During the pilot study we discussed this possibility with users [8]. They said that they are interested in such a feature. While some users would like to practise with situations that are close to them, others noted that they would like it the other way around as that would be less personal. Moreover, users also noted that it could be beneficial to learn how to generalise the application of self-compassion, and thus it would be good to also practise with more unfamiliar situations.

Thus, the goal is to create a topic modelling algorithm that can be used to determine which topics are discussed in freely written, user created, content, which can then be used to personalise the exercises of the training. As we use predetermined topics, we cannot use existing algorithms such as those mentioned in Sect. 2. With data from the pilot study we can analyse and evaluate how the algorithm performs and how it can be embedded in the training website.

3.1 Data

In the previously mentioned pilot study, 24 users worked with the website for a limited period of time [8]. All participants were female, with a mean age of 22.5 (SD = 2.04). From these 24 users, 18 users actually entered data in the website. Of these 18 users, the data entered in the exercises and journals is saved. For this study we use the data from the gratitude and self-compassion journal, and one of the exercises called 'Practise self-compassion'. These contain texts about situations from the users' daily lives. All data is in Dutch.

In the gratitude journal users are asked to fill in something they are grateful for on that day, and characterise their answer with a short tag (max. 50 characters). Both this description as well as the tag can be used to determine the topic of the text. For the self-compassion journal users describe a situation, characterise this with a similar short tag and explain how self-compassion is used or could be used in that situation. From this journal we use the situation description and the tag. Finally, in the exercise users are asked if there is something they are struggling with and to describe this situation. Even if they are not struggling at the moment, they are asked to use a situation from their (recent) past. This description is used for the topic modelling. For the exercise no tag is available. In total, the available data contains 181 unique notes: 61 gratitude journal notes, 54 self-compassion journal notes, and 66 exercises. Not all participants contributed the same number of notes. Figure 1 shows how many notes each participant contributed.

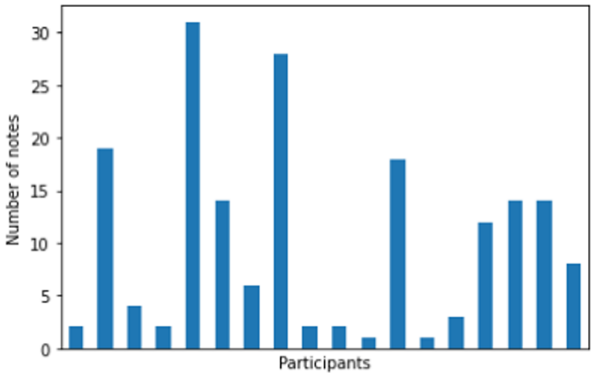

Fig. 1. Number of notes per participant

3.2 Topics and Data Labelling

In the exercises, situations around the topics of *friendship*, *body insecurity*, *social anxiety*, and *social media* are used. The choice for these topics was related to the aim of the training. However, based on the pilot study it became clear that other topics were missing [8]. The data from the interviews of the pilot study can be used to find more suitable topics.

In addition, we analysed the data to see which topics are actually discussed by the users. To do so, the data was first labelled using the four predefined labels. If none of the labels suited the text, 'no topic' was noted. When the tag matched a label, this label was noted even though the text might be less clear. When multiple topics could apply, the label that the majority of the text applied to was noted. If the text was divided equally, 'no topic' was noted. The three researchers checked the labelling and discussed any discrepancies. The labels given by the researchers are called the gold labels. After this, we studied the notes that had no label, and decided on a new set of labels based on both the data findings as well as the results from the interviews. We changed *friendship* into *relationships*, as family or romantic relationships were also discussed. We changed *body insecurity* to *body image* as this suits better with positive message about your body. The new set of labels is: *social anxiety, social media, relationships, body image, school, job,* and *emotions*. Where emotions is a topic that covers all texts where emotions, moods or feelings have an important role. Although *social media* was not used by any of the participants, we do keep it as a topic as there is already content on the website with that topic.

3.3 Algorithm

For all predefined topics, related words were gathered using the Related Words website[1] that gives a list of words that are related to a given search term. We

[1] https://relatedwords.org/.

created our related word lists based on the terms: 'social media', 'social anxiety', 'job', and 'emotions' for the eponymous topics, and 'body' for the *body image* topic, 'relationship' and 'friend' for the *relationship* topic, and 'school' and 'test' for the *school* topic.

For the related word lists we first included the topic name itself. Next, we excluded words from the related word list that are clearly unrelated to our topic, such as names of people or events, drugs and disorders. For most topics we looked only at the first 50 terms. For the term 'friend', we only looked at words for friends or family members. For the term 'test', we only looked at terms related to grading in the school context. For the term 'social media' we included some currently popular social media platforms that are not included yet on the Related Words website. The related word lists contain 51 words for *social media*, 62 words for *body image*, 32 words for *social anxiety*, 116 words for *relationships*, 52 words for *job*, 88 words for *school*, and 51 words for *emotions*.

Data Preprocessing. Before analysing the data, we remove any names mentioned in the data to ensure anonimity. In most cases, the names are replaced by an X to make the text still readable for the labelling process. The texts and tags are translated to English with the Google Translate API[2]. These English texts and tags are used for the further preprocessing.

The texts are preprocessed with the help of the Natural Language Toolkit (NLTK) for Python[3]. To do this, the text is first tokenized. With this tokenization, the text is separated into words (tokens). All remaining tokens with more then two characters are saved, other tokens are removed as they are not valuable. Using the NLTK stopwords list we remove stopwords from the remaining tokens. Stopwords are words that are frequently used in human language. They are removed because they often do not add much value to a sentence. The final step is lemmatizing the words and verbs. This means that verbs are turned into their present tense and plural words are put in their singular form.

The tags in the data and the related word lists are preprocessed the same way. Due to this preprocessing, some tags might be deleted. In the preprocessing of the related words we also remove words that appear multiple times for the same topic. This can happen when multiple words included the same term, and after splitting the terms in single words these duplicate words are removed.

Topic Modelling. The algorithm that is developed looks at the overlap between the prepared text and tag and the prepared related words of each topic. First, it counts the number of related words that are present in the prepared text. Each overlap adds one point to the similarity score of the topic. When a word is used multiple times, each occurance is counted. Next, it is checked whether the related words overlap with the tag (if present). If this is the case, 3 points are added to the similarity score. Finally, the topic label itself is tokenized and

[2] https://pypi.org/project/google-trans-new/.
[3] https://www.nltk.org/.

it is checked whether the words from that tokenized label overlap with the tag. If this is the case, the similarity score is increased with 6 points.

The higher increments for an overlap with the tag are based on the fact that the tag is the shortest description of the note, so in general it has a higher likelihood of being a description of the topic of the text. When the tag overlaps with the name of the topic, this likelihood is even higher.

Once the similarity score is determined for each topic, the predicted topic is chosen. For this, we first check if the highest similarity score is higher than a threshold value. During the evaluation we will choose the right threshold value. If the highest score is above the threshold value, and this similarity score is only calculated for one topic, this topic is predicted. In other cases, the algorithm cannot be sure about the topic and thus predicts multiple topics. When evaluating the algorithm we look at the combinations of topics that are predicted and their gold labels. If we can find patterns in this, these patterns can be used to make rules about the final prediction of the algorithm. If not, the algorithm will predict 'no topic' when multiple topics score equally.

The algorithm needs to predict at most one topic for every text written by users for one of the data sources mentioned in Sect. 3.1. This topic will be saved for that text, but can later be manually changed by the user (choosing from the predefined set of topics).

3.4 Algorithm Evaluation

To analyse the algorithm, we look at the accuracy with which it predicts the labels. To calculate this, the number of correctly predicted labels is divided by the total number of predictions. The higher the accuracy, the better. However, this accuracy is not everything. We also need to look at what goes wrong. In our application we consider it less problematic if the algorithm predicts 'no topic' instead of a wrong topic. Users will be able to manually add or edit a topic labelling if they want. However, if they do not correct mistakes, it is better if mistakes are avoided to prevent wrong displays of frequently used topics. Thus, we also calculate the accuracy of correct predictions and 'no topic' predictions compared to the total number of predictions. However, this accuracy cannot be used to improve the model, as that would mean that a 100% accuracy could be achieved by simply always predicting 'no topic'. Therefore we use the first accuracy to determine the best threshold for the points to be considered a labelling and to study the multiple label predictions to determine the final predictions.

4 Results

4.1 Data Analysis

As shown in Table 1, most of the entries have no label. Furthermore, it can be seen that *social media* has not been used as a topic in any of the data. In the gratitude journal only *friendship* is discussed, which is the topic with the most

Table 1. Numbers of notes labelled with initial labels

	Social anxiety	Social media	Friendship	Body insecurity	No topic
Gratitude	0	0	16	0	45
Journal	3	0	2	4	45
Exercise	4	0	5	5	52
Total	*7*	*0*	*23*	*9*	*142*

Table 2. Numbers of notes labelled with extended labels

	Social anxiety	Social media	Relationships	Body image	Job	School	Emotions	No topic
Gratitude	0	0	24	2	3	4	1	27
Journal	3	0	6	5	8	11	13	8
Exercise	4	0	6	5	5	10	26	10
Total	*7*	*0*	*36*	*12*	*16*	*25*	*40*	*45*

positive wording. Table 2 shows the number of counted labels when the extended set of labels is used. Still, a quarter of the texts is not labelled.

Most of the notes without a label are small notes. Figure 2 shows that the number of texts without a tag reduces if you use a threshold for the minimum number of words in a text (without preprocessing). When using a threshold of 10 the number of unlabeled texts halves, when using a threshold of 20 it reduces to 29% of the original number of unlabeld texts and with a threshold of 30 this is only 18%. However, the other line shows that the total number of texts also reduces, at a higher rate than the unlabeled data. To include as much data as possible, we will use all notes. However, we will compare it with different thresholds to see if this effects the accuracies.

We counted the words appearing in the prepared texts of the different topics, and looked at the words that were used most frequently and more than five times. As there are no texts classified for *social media*, also no words could be found. Table 3 shows that most words are clearly unrelated to the topic, such as 'n't' or 'good'. However, the word 'colleague' makes sense for *job* as well as the word 'felt' for *emotions*. 'Felt' should have been changed to 'feel' in the lemmatization. However, as 'felt' is also a noun, it was not recognized as a verb. Both 'felt' and 'colleague' are added to the related word list.

4.2 Results on Topic Modelling

First, we predict topics for each note with a points threshold of 1. If multiple topics have the highest score, we predict them all. Now we can see if there can be a pattern found in the predictions of multiple topics and their gold label. When looking at the combinations, such a pattern cannot be found. Thus, we choose that if multiple labels are predicted, the predicted label is 'no topic'.

Table 3. Gold labels and words that appear >5 times, * included in related word list

Gold label	Words
Body image	good (n = 6)
Social Anxiety	n't (n = 6)
Relationships	grateful (n = 14), good (n = 8), girlfriend* (n = 8), nice (n = 7), n't (n = 8), feel (n = 8)
Job	today (n = 7), colleague (n = 7), work* (n = 12), feel (n = 10), job* (n = 6), would (n = 7), make (n = 6), mistake (n = 6), n't (n = 7), say (n = 6), message (n = 6)
School	today (n = 8), school* (n = 7), lot (n = 6), exam* (n = 7), n't (n = 6)
Emotions	could (n = 7), feel* (n = 17), felt (n = 7), good (n = 8), get (n = 7), n't (n = 6)

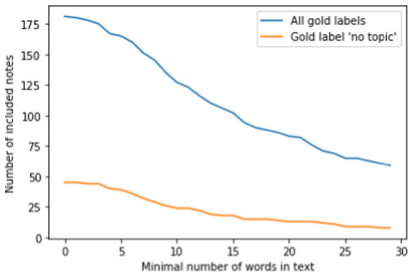

Fig. 2. Number of notes when using a minimum number of words

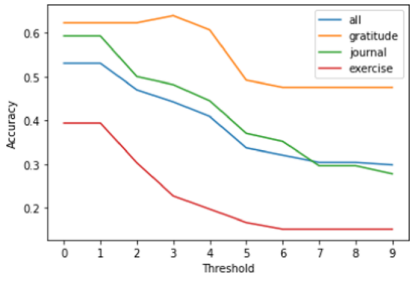

Fig. 3. Accuracies for predicted labels using threshold 1–10

Next, we test thresholds 1-10 to see which threshold has the best accuracy when using the determined prediction rule for multiple labels. From Fig. 3 it is clear that a threshold of 1 provides the best overall accuracy. Table 4 shows the accuracies that are predicted with this 1-point threshold. Both for the gratitude and journal notes the accuracy is higher than the average accuracy, but for the exercises this is lower. A difference between the gratitude/journal notes and the exercises is that the exercises do not have a tag. The accuracy of the 106 notes with a tag is 62.3%, the accuracy of the 75 notes without a tag[4] is 42.7%. Based on a Student's t-test we can conclude that this difference is significant (p-value = 0.0090).

We also analysed if using a minimum of words in the original text has an effect on the accuracy, which can be found in Table 4 as well. Overall, this does not seem to have an effect as most of the accuracies are similar. Only for gratitude there is a big difference when using only notes that have >20 words. This category also loses 65.6% of its notes when using only >20 words.

[4] 66 exercises notes + 9 gratitude/journal notes where the tag is removed in the preparation process

Table 4. Accuracies of different components with different word thresholds. Accuracy in brackets is accuracy including 'no topic'-predictions as correct predictions

Accuracy	All (n = 181)	>10 words (n = 127)	>20 words (n = 83)	>30 words (n = 55)
All	54.1% (80.4%)	54.3% (79.5%)	59.0% (77.1%)	58.2% (74.5%)
Gratitude	62.3% (77.0%)	69.4% (77.8%)	80.9% (85.7%)	63.5% (72.7%)
Journal	61.1% (85.2%)	62.5% (85.0%)	58.6% (82.8%)	54.5% (81.8%)
Exercises	40.9% (78.8%)	37.2% (76.5%)	45.4% (66.7%)	59.1% (68.2%)

Table 5 shows the number of notes for each topic that did not get the right prediction. The highest number of incorrect predictions are made for notes with the gold label *emotions*. However, this is also the biggest category. Relatively, the most incorrect predictions are made for gold label *social anxiety*.

Table 5. Incorrect predictions for each gold label

Gold label	Including 'no topic'	Excluding 'no topic'
Emotions	27 (67.5%)	10 (25.0%)
Relationships	20 (55.5%)	5 (13.9%)
No topic	9 (20.0%)	9 (20.0%)
Body image	8 (66.7%)	3 (25%)
School	8 (32.0%)	4 (16.0%)
Social anxiety	6 (85.7%)	3 (42.9%)
Job	5 (31.2%)	2 (12.5%)

As explained in Sect. 3.4, in our application it is better when the algorithm predicts 'no topic' instead of a wrong topic. Therefore we also calculated the accuracy when including the 'no topic'-predictions as correct predictions, again using point threshold 1. The accuracies are shown in brackets in Table 4. In Table 5 it can be seen that still the most incorrect predictions are made for gold label *emotions*, but also for *no topic*. Relatively, *social anxiety* has the most incorrect predictions.

5 Conclusion and Discussion

The goal of this paper is to explore the possibility of using topic modelling to personalise the experience of users of our self-compassion training website. Based on the accuracy of the model it seems that to some extent it is possible to predict the topic for different texts from users. Especially when you consider a 'no topic'-prediction as a correct prediction as well. This makes sense in our application as users would be able to change or add the label manually afterwards.

For our topic modelling algorithm we choose to work with a similarity score for words and related words. Other approaches would be to use machine learning. The analysis of the data showed that on word level there are hardly any words that are characteristic for specific topics. It is thus unclear if using such approaches would make sense.

We observe that the preprocessed data loses meaning. Often the topic of a text is found by the human reader in the combination of sentences and wordings. For example a sentence like 'I have a bad headache and my shoulders are stiff due to my stress about my upcoming exam.' will be preprocessed into the words: 'bad', 'headache', 'shoulder', 'stiff', 'due', 'stress', 'upcoming', and 'exam'. Based on only this text, the algorithm will predict both the topics *school* and *body image*, as it includes the word 'shoulder' and 'exam'. However, when only looking at the prepared text it is also harder for the human reader to decide what topic this text is about. It is thus interesting to explore whether a different preprocessing could help the algorithm to perform better. However, as the data is divided over many different topics, the remaining data is limited in its size and thus it is hard to draw conclusions on this. Another note that needs to be made, is that little changes in the preprocessing could have an effect on the accuracy of the algorithm. For translating, we use an API. If something in this API changes, the texts and thus the outcomes could be different. Also, spelling errors or ambiguous words effect the outcomes of the algorithm.

Participants of the pilot were not aware of any of the topics, they were completely free when writing their notes. The data therefore is very comparable to how it will be in the actual evaluation study of our training website. This holds for the content of the data, but also for the form (length of texts, writing errors etcetera). The only difference might be that in the new version of the training the minimum length of texts will be increased based on the average lengths found in the pilot data. We could have used more data, for example from our database of situation texts, but as this data is written with the topics in mind this would not be representable for the texts of the users.

With this simple form of topic modelling we can make sure that to some degree there can be personalization in the training. We plan to include a word cloud with the topics that have been predicted by the algorithm. Moreover, we will use it to ask users if they want to practise with a situation close to them (from one of their frequently discussed topics) or something less personal (from one of the topics they do not discuss (often)). In conclusion, with the topic modelling algorithms described in this paper, it becomes possible to personalise parts of the self-compassion training that the algorithm is developed for. With the proposed uses of this, the user-experience of the training will be increased.

References

1. Berry, M.W., Mohamed, A., Yap, B.W.: Supervised and Unsupervised Learning for Data Science. Springer, Heidelberg (2019). https://doi.org/10.1007/978-3-030-22475-2
2. Blei, D.M.: Probabilistic topic models. Commun. ACM **55**(4), 77–84 (2012)

3. Ferrari, M., Hunt, C., Harrysunker, A., Abbott, M.J., Beath, A.P., Einstein, D.A.: Self-compassion interventions and psychosocial outcomes: a meta-analysis of RCTs. Mindfulness **10**(8), 1455–1473 (2019)
4. Finlay-Jones, A., Kane, R., Rees, C.: Self-compassion online: a pilot study of an internet-based self-compassion cultivation program for psychology trainees. J. Clin. Psychol. **73**(7), 797–816 (2017)
5. Gulliver, A., Griffiths, K.M., Christensen, H.: Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry **10**(1), 1–9 (2010)
6. Kherwa, P., Bansal, P.: Topic modeling: a comprehensive review. EAI Endors. Trans. Scalable Inf. Syst. **7**(24) (2020)
7. van der Lubbe, L.M., Gerritsen, C., Klein, M.C., Hindriks, K.V., Rodgers, R.F.: Designing a gamified self-compassion training. In: 22st Annual European GAMEON® Conference: Simulation and AI in Computer Games (2021)
8. van der Lubbe, L.M., Gerritsen, C., Klein, M.C., Hindriks, K.V., Rodgers, R.F.: A pilot study of a gamified self-compassion training. In: 22st Annual European GAMEON® Conference: Simulation and AI in Computer Games (2021)
9. Neff, K.D., Rude, S.S., Kirkpatrick, K.L.: An examination of self-compassion in relation to positive psychological functioning and personality traits. J. Res. Pers. **41**(4), 908–916 (2007)
10. Talbot, F., Thériault, J., French, D.J.: Self-compassion: evaluation of a psychoeducational website. Behav. Cogn. Psychother. **45**(2), 198 (2017)
11. Ustün, T.: The global burden of mental disorders. Am. J. Public Health **89**(9), 1315–1318 (1999)