**Understanding volunteering intentions with computational text analysis**

Elisabet Doodeman (e.g.m.doodeman@student.vu.nl) VU Amsterdam

Arjen De Wit (a.de.wit@vu.nl ) VU Amsterdam)

John Mohan (j.mohan@bham.ac.uk) TSRC, University of Birmingham

**1. Introduction**

There is a wealth of empirical work on different forms of volunteering, but criticisms have been raised about the operationalization of concepts such as volunteering in research studies. The main argument being made here is that conceptualizations said to encapsulate voluntary action are grounded in formal, traditional, and static interpretations of voluntary action – not least because the questions asked about volunteering are restrictive, and presume a shared understanding of volunteering on the part of those responding to them (Steinberg, Rooney & Chin, 2016; Wiepking, 2021).

 Furthermore, such work has generally focussed on the characteristics of the active volunteer force in the specific time period of the study itself (Musick & Wilson, 2018). Thus we know much about particular types of citizen: those who identify themselves with the conceptualizations discussed above. Studies have shown that this subset represents a relatively small group of people that tends to be (already) highly active in the voluntary sector. A corollary is that a possibly larger group of contributors remains invisible to us. Firstly, there are those who are expressing behaviours that can be seen as a contribution to the sector, yet do not fit into the classical volunteer profile or even identify as such. Missing out on their contribution could further skew the picture we have painted at this point. A similar argument could be made for the absence of any real overview of *potential* volunteers, yet understanding this group and their motives to engage with the sector could be the key to unlocking a great potential force. Some time ago now, Tonkens (2010) referred to these two groups as one ‘altruistic surplus’ that was ought to be recognized and exploited accordingly.

A third issue we would like to draw attention toward is rooted in the general nature of previous studies, which have come under scrutiny for largely ignoring cultural and contextual factors (Hustinx et al., 2022). Questions surrounding different interpretations and biographical factors are coming to the forefront (Shachar et al., 2019). The conceptual issues presented above can be largely attributed to the research methods of these studies. Data collection has mostly involved large-scale survey research which requires individuals to respond to a menu of pre-set conceptualizations, which may limit the exploration of complex contextual factors. Of course we do have qualitative studies of the meaning of volunteering and its place in people’s lives, but these are small scale and this raises questions about the generalisability of findings.

In order to address the issues presented above, we have attempted to capture a dimension of the attitudes of citizens to volunteering, namely whether such pro-social actions feature spontaneously in descriptions by individuals of their imagined future lives. We opted for a study design that would embrace the benefits of both quantitative and qualitative research methods while avoiding the pitfall described before. This initial paper will lay out the process in phases, from the start to the current state of our project.

**2. Background**

In the UK, much faith has been invested in the idea that as prosperous baby-boomers retired they would become active in their communities. In a 1981 speech the then Prime Minister Margaret Thatcher anticipated that there would be “many more retired people, hale and hearty, wanting to do something to help others.” (Thatcher, 1981). Some twenty years later, volunteering rates hadn’t changed much, and the Conservative MP Oliver Letwin was to be heard expressing the hope, nevertheless, that increased longevity meant that a “huge new well” of active, healthy, energetic, experienced, and intelligent people who were retired or approaching it would be available; what was needed was a way of tapping that resource.

Letwin didn’t develop ideas on how that well might have been tapped. Unlike a natural resource which can be extracted, gushing forth as the drill breaks into the subterranean reservoir, volunteers do not automatically come when they are summoned. In recent decades optimistic commentators have nevertheless anticipated a sustained increase in volunteering partly because of favourable demographics and partly because of a sense that people wanted to be part of something bigger than themselves – whether this be major sporting events or national celebrations (Commonwealth and Olympic Games; most recently, the coronation of King Charles) in which volunteers have played prominent roles. Others pointed to the compassionate response to support neighbours and front-line workers in the Covid emergency as something on which to build. Yet it is incontestable that volunteering in the UK has not increased, despite all the attention given to it. Why might this be so?

Answers might be found in a combination of social attitudes and economic circumstances and it’s the former that we discuss in this paper. We know relatively little about attitudes to volunteering. The small number of robust UK surveys on this topic are now some thirty years old, but are revealing about individuals’ perceptions as to the necessity and desirability of voluntary action – albeit at a specific time period, the years of John Major’s premiership, a period of Conservative disharmony and also disinvestment from public services (a period which therefore has much in common with the present day). Nor do we know very much about future *intentions* regarding voluntary action. Those claiming a sudden upturn in volunteering have been known to overplay their hand. In the UK a prominent think tank closely associated with the new government proclaimed that volunteering rates had allegedly “doubled” in a short time period. The evidence base for this appeared to have been newspaper reports which suggested a relationship between an increase in the demand for volunteering opportunities and the rise in unemployment induced by the Great Recession of 2008-9. A more recent Pro Bono Economics study (2023) claimed that there is substantial evidence of pent-up demand to become engaged in volunteering, and the authors projected an increase of several million volunteers on the back of a survey which stated that a quarter of adults *intended to volunteer* during 2023.[3] If shown to be true, this would be a completely unprecedented increase, but as far as can be judged the survey didn’t distinguish those who intended to continue volunteering from those who were contemplating doing so for the first time (a proportion actually likely to be quite low, because people move into and out of volunteering during their lives, and therefore not all will be entirely new to volunteering).

The problem is the reliability of such statements of intentions. Volunteering is a socially desirable activity, and we’re naturally predisposed to give the responses to questions that we think will elicit approbation. But we shouldn’t believe that such responses are a sure guide to how people will actually behave. Attitudes and behaviours are also framed by context; people do not act in a vacuum. The economic and social context influences the resources that people possess and the social networks they have but appeals to volunteer are framed very much against particular circumstances. There is no way of telling whether the Big Society project of David Cameron would have received a more positive reception had it been launched in more auspicious times, but it was widely perceived as a cover for substantial cuts in public expenditure, and attitudes were therefore negative. (Lindsey and Bulloch, 2013).

However, while the UK possesses an exceptional time series of data on social attitudes, only rarely have questions been posed about attitudes to volunteering. National surveys of volunteering conducted in 1981 and 1991 asked respondents were asked to react to propositions that “a society with voluntary workers shows that it is a caring society” and that voluntary workers “offer something different that can never be provided by the state”. More specific questions on voluntary action were asked by the British Social Attitudes Survey (BSAS), probing the extent of agreement with statements about whether or not they agreed that “everyone has a duty to do voluntary work”, whether “as a society, we rely too much on volunteers”, and whether “voluntary work is a good thing for volunteers because it makes them feel they are contributing to society”. These survey findings are interesting particularly in relation to society’s reliance on volunteers and the question of whether there ought to be a duty to volunteer.

UK attitudinal surveys have also explored views about the relative roles of statutory and private responsibility for welfare. The British data demonstrates a cyclical pattern of attitudes either for or against greater government intervention, self-reliance, and higher welfare payments. Generally Support for government intervention rose and fell, broadly in inverse relationship to the state of the economy and the paucity of government support for those in need. While attitudes have at times swung away from support for state intervention, there has been no shift in levels of voluntary action; behaviour was not switching in favour of greater voluntarism by individuals. Unfortunately, we do not have survey data on which we might base further discussion – for example, more comprehensive data on attitudes to voluntarism – which would enable us to explore the relationships between attitudes and pro-social behaviours in more depth. Importantly, none of these surveys, which indicate broadly supportive attitudes towards voluntary action, contain any information about whether people felt they were likely to volunteer in the future, or not.

One way into this subject is a module of questions issued to a subset of participants in the BHPS / Understanding Society longitudinal surveys, namely people aged over 45 who were not retired, and those of pensionable age (this was 65 in the UK but was raised to 66 from 2020), aged under 71, who did not consider themselves to be retired. They were offered “a list of things that some people say are important about retirement” - being one’s own boss, having time to take things easy, the opportunity to travel, time for family and friends, time for leisure, and time for voluntary work – and asked to say whether these were very important, moderately important, somewhat important, or not important.

Having time for voluntary work was the least popular of these options – around 12% regarded it as “very” important and 28% rated it of “moderate” importance. In contrast, between two-thirds and over four fifths of respondents rated the options of relaxation, travel, time for friends and family, and leisure pursuits as either very or moderately important, with over 55% indicating that family and friends were very important to them. The relative rankings and proportions involved have been broadly constant over time – voluntary work is always the response that attracts the least positive endorsement, while when these questions were first posed in 2001, 11% and 27% indicated that time for voluntary work was very and moderately important, respectively. Even among those who were already active volunteers, voluntary work was the least popular choice. This gives us a broad ordering of people’s priorities and clearly indicates that volunteering is by no means uppermost in people’s minds when compared with other potential priorities.

**3. Data and Descriptives**

How can we move beyond the frequency counts and rankings of survey data? We now describe a source and method which to our knowledge has not previously been deployed in this field. Our data were drawn from the tenth sweep of the National Child Development Survey (2008). The NCDS is currently conducted by the Centre for Longitudinal Studies as part of a wider suite of British longitudinal surveys (see <https://www.closer.ac.uk>). The study, which has been collecting data since the birth of its 12000 initial participants (all those born in a selected week in 1958), aims to chart various social, economic, and health conditions throughout their lifetime. Cohort members are re-surveyed at intervals – usually every seven years, though with some gaps. The full survey design and data can be publicly accessed (UK Data service, 2022). Though there is attrition due to death, loss of contact with participants, and unwillingness to engage with individual survey waves, there were still 9760 cohort members who participated in the 2008 sweep, when they were aged 50 – a wave of the study of particular interest to us. From time to time this study has offered participants the opportunity to respond in writing to open-ended questions, in whatever way they saw fit. Thus the survey provides extensive quantitative data on people’s life courses, complemented by rich, textual descriptive information.

Our immediate attention was drawn to the final question from the 2008 survey. In this essay-type question, cohort members were asked to imagine their life in ten years. They were asked to describe their thoughts, with specific attention to their interests, home life, health and well-being, and any work they might be doing. The wording of the prompt was:

“Imagine that you are now 60 years old. Please write a few lines about the life you are leading (your interests, your home life, your health and well-being, and any work you may be doing)”. (Elliott, 2020).

We expected responses to indicate a detailed contextual picture of the lives people expect to live in a certain time period. Although not specifically related to voluntary action, we expected the data to hold descriptive information on a variety of behaviours that could – we hoped - include in-depth insights on (1) currently invisible voluntary action and (2) future potential voluntary action.

What is especially interesting about the timing of this question, is that when members were asked about this, they would likely have expected to approach retirement age in or after those 10 years. As it is often assumed that this Baby Boom Generation will enter the voluntary sector at this time, this data could hold rather important information for testing the assumption that they would be actively considering becoming involved in voluntary action. In total, 7378 members of this survey cohort responded, although in many cases only a few words or sentences were written. An initial and basic analysis, simply using frequency counts of common terms such as “voluntary work”, “volunteering”, or “charity work”, showed relatively few direct references to anything that was recognisable as voluntary action. There were 450 responses including the phrase “voluntary work”, and another 55 featuring the word "volunteer", while 141 referred to "helping". However, quick inspection of the latter shows a number were referring to helping children and grandchildren, so that is not an unambiguous indication (and of course this relates to wider confusions about what counts as voluntary action, as well as providing an indication of respondent priorities). A further 76 referred to charities including one who envisages that they will "work two days a week in the local charity shop". Differences in phraseology are fascinating: some report “helping” in the charity shop while others characterise their prospective activity as “working”; another stated that in their future existence they would “volunteer for a couple of charities and help at the hospice" – raising the question what is the difference between volunteering and helping. Another describes "taking part in charity activities, [and] helping other people in my community" – perhaps echoing the lack of clarity about what counts as formal and informal volunteering.

Approximately 11% made some reference to activities of this kind, which was around half the proportion (typically around 23%) which reported involvement in “unpaid voluntary work” in the same wave of this survey. What was more noticeable than responses to the closed survey questions was the differences in responses by men and women – the latter being nearly 3 times as likely as men to make some reference to community involvement, a disparity far greater than that recorded in the survey question on unpaid voluntary work. There was a similar gradient for education – ranging from 17% for those with a degree or higher degree, to 7% for those who had left school with very few educational qualifications. We would anticipate higher levels of education to be apparent in responses such as these.

There was also a clear gradient in the responses between those who were already volunteering and those who weren’t. Of those cohort members who were volunteering at least once a week in 2008, at least 15% said that they would be engaged in volunteering at age 60, compared to less than half of that for those volunteering once a year or less. Perhaps surprisingly as many as 4% of those volunteering “never or almost never” at the age of 50 indicated that they might be volunteering when they reached 60.

These frequency counts don’t allow much beyond a description of who spontaneously expresses interest in some sort of voluntary action. Further analysis was therefore required to illuminate where volunteering is placed in relation to individuals’ priorities, and to gain more insight into what exactly it was that respondents thought they’d be doing ten years hence, and that required a different approach.

**5. (Introduction to) Computational Text Analysis**

After acquiring this rich and voluminous data set, we had to consider how to go about analysing it. We turned to a relatively more novel (to this field) set of methods. Through the use of so-called computational text analysis methods, large amounts of textual data can be structured in a way that is meaningful to researchers (Jacobi, van Atteveldt & Welbers, 2016). Rather than involving much manual labour, a computer performs most of this task. Researchers' function lies in the interpretation of the computational output. In this way, both data quality and quantity can remain intact and add their relative strengths to a study.

*Data and previous work*

Interestingly, such methods have been applied to our data set before. With regard to our specific survey question, two recent studies are relevant. To our knowledge, Elliot (2012) was the first to work with this specific sample. She published a preliminary descriptive analysis of a sub-sample (n=500) of individuals that answered this question. The aim here was to provide a broad overview of answers in general. By sampling, Elliot did compromise on data quantity. Moreover, as the analysis was done in NVivo(9), computation analysis could only go as far as some explorative world frequency output. Table 1 lays out the most frequent words in the dataset (as presented by Elliot in their paper). Through this structured format, we could get a slight overview of the themes we were likely to encounter.



Table 1: word frequencies as presented by Elliot (2012)

Weber (2021) used a computational topic model to compare imagined futures as outlined by different genders and social classes. Modelling was done in the programming software R Studio, which offers many more options than NVivo. Where NVivo is generally used to manually code qualitative studies, R Studio is a programming language, that leans more toward the quantitative side. As such a computational process starts out purely quantitatively, this is a better fit for such a task. Table 2 displays the topics that Weber presented in their paper.

Table 2: Topic Model as produced by Weber (2021)

*Considerations*

The studies outlined above are relatively interesting formats for our purpose. However, they have not had the same focus as we did. As can be observed from the tables above, both studies provided a broad and general overview of the nature of answers. From the work of Elliot, we now know a) what words people tend to use to describe similar concepts and b) that put together they seemed to cover multiple facets of life. As a result, we felt more secure in the richness of our dataset.

What Weber showed us was how these little pieces of text contained information on a variety of lifestyle choices. Moreover, Weber’s work showed how these choices appeared to have a relationship with an individual’s socioeconomic circumstances. Factors concerning personal values and interests may not be directly related to voluntary action but offer insight into the frameworks people use to manage the time they spend on certain activities. The consideration of these frames could lead to insights into possible future contributions an individual might make to volunteering. Further, these representations could hold information on what is already being done in other domains (e.g., family) that can be considered voluntary contributions in themselves or perhaps barriers to such contributions. For our purpose, we hoped to distil the information that responses hold concerning such broader voluntary actions and intentions.

*Related Work*

When we turn to literature closer to our topic of interest, some experimental studies can be found. An example is De Wit (2020), who experimented with a topic modelling technique to analyse open-ended survey questions, drawn from the A Broader Mind Longitudinal Survey. This survey also included an open question quite similar to ours in the sense that participants were asked to imagine and describe what their lives would look like in a future timeframe. Otherwise, the question was formulated a bit more specifically than the one we were dealing with (geared toward career and positive emotions). Topic modelling output proved quite insightful on its own, but could also be used as input for ANOVA analysis. This analysis was done on a subset (n=200) of the data set but appears promising for larger quantities of data.

De Wit was able to distinguish three topics within the answers given to the open question that reflected the importance attached by respondents to a) a steady job, b) sufficient time for family, and c) a contribution to society. Interestingly, it was also found that question formulation tended to affect the answers that were given. E.g., when asked about retirement specifically, the likelihood of answers covering the importance of a steady job decreased, while the importance of a contribution to society increased. Given that our question does not specifically ask about retirement, yet could be considered as associated with retirement, we do not have a strong prediction here. We will return to the consequences of question formulation later when we discuss our own findings.

**5. The Methodological Process/Exploration of Methods/Methodological Approach...**

**5.1 LDA Topic Modelling**

Following in the footsteps of these examples, we have set out to optimise the potential of the NCDS dataset. In the following sections, we will present an explorative study, aiming to use computational text analysis to structure the data set as distilled from our survey question of interest. first, we introduce our chosen approach. Topic modelling has been described as ‘one of the most powerful techniques in text mining for data mining, latent data discovery, and finding relationships among data and text documents’ (Jelodar, 2019). It is a type of computational text analysis method that can identify semantic structures in large volumes of text. This model produces *topics*, which are essentially clusters of words that are close in similarity. Words are automatically assigned to topics, without human supervision, but researchers are involved in the final product as they set the model's parameters and interpreting its outputs. Also, topic modelling does not provide any guidance on the interpretation of topics, which is up to the researchers again.

Several social scientists have experimented with topic modelling when aiming to handle large volumes of data. Latent Dirichlet Allocation (LDA) is often referred to as the most popular topic model approach for natural language processing. The coding of the model is easily facilitated through the open-source statistical package R.

**5.2 LDA and NCDS**

Here the process and performance of our LDA topic model are discussed. We have followed general guidelines as set by Jacobi, Van Atteveldt and Welbers (2016).

*Data Preparation*

Topic modelling operates in an unsupervised way. During the process, a collection of documents is broken down into clusters of words. To avoid any confusion for the researcher in the interpretation stage, data cleaning is important. This can be done through low-level string operations. For us, this meant the removal of certain symbols that likely resulted from an encoding issue and/or were related to the transcription style. E.g., ‘@’ was used a lot to refer to certain dimensions of the question. This might have given some type of overview for manual coding but could disturb the LDA process due to its omnipresence and relatively unimportant function to us. Next, empty rows and columns were deleted. We then proceeded to select our columns of interest: the textual data and NCDS respondent numbers. Lastly, we added row numbers to our rows to be able to match data to topics later on.

*Pre-processing*

During the pre-processing stage, raw language data is processed into representations that computers can work with in terms of calculations. For this purpose, a document term matrix was created from our corpus (textual data). In this way, the presence of certain words in the documents was documented. The process of breaking down the text into smaller components (mostly words) is called tokenization. To create the most elegant version of the document term matrix, we also applied functions that ensured lowercasing, stemming, and the removal of stop words. The first two processes create a more efficient matrix by transforming all capitals to lower cases and by breaking down certain words (verbs) to their stem. As a result, the same or very similar expressions are now grouped together and recognized as linked by the model. Stop words were removed as they are not expected to be very relevant for our analysis, while they might clog the topic model and lead to outputs that are hard to interpret on the behalf of the researcher.

After pre-processing our matrix was converted to a different format that was appropriate for topic modelling. Content wise, these matrices are the same. A sample seed was set to ensure the reproducibility of results.

*Parameter Settings*

The parameter settings are the first point of human intervention. Firstly, researchers are required to manually specify the number of topics for the model. Ideally, we would want to establish a model with the lowest number of topics without losing relevant information. Unfortunately, there is no default option for this, nor a simple rule of thumb. However, certain approaches could guide this decision-making process.

For example, researchers could compare the output of several models (a variety of topics) based on face validity and construct validity. Both are highest for the model that 1) contains terms that make the most sense to the researcher and 2) has the most cohesive and distinctive topics. The model with the highest validity levels is then considered the most appropriate.

An issue one might run into here is the question of how many models should be compared. While human evaluation of all possible models could produce trustworthy results, this is obviously time-consuming and costly . Luckily, there are ways to work around this issue through the guidance of quantitative metrics. Because these topic models are probabilistic it is possible to calculate how well the models tend to function. To do so, researchers have to train an LDA model on only a portion of the data. The other data is used for model evaluation. This routine is repeated on a large number of models (with varying topic numbers or k). The output of such a test is presented through a perplexity plot. Much like a goodness-of-fit measure, the model that best predicts the data is considered a good indicator for the right number of topics. Perplexity can be defined as the normalised inverse probability of the test set. Therefore, lower perplexity indicates better function according to automatic evaluation.



Figure 1: Perplexity Plot

We trained our model on 75% of our data. Up until 30 topics, perplexity decreases as the number of topics increases. After this point, perplexity increases alongside the increasing topics. From looking at this plot, we can estimate the ideal number of topics to be between 15 and 30 topics. The 30-topic option seems most suitable in terms of scores but can be considered quite a lot of topics for manual interpretation. Several topic settings were explored within this range. Lastly, we considered alpha. Alpha is a hyperparameter controlling per-document topic distribution. The higher the alpha score, the more likely it becomes for documents to contain a mixture of topics. In our case, alpha was set to k/50, which is in line with field standards.

*Interpretation*

When evaluating and interpreting these selected models via human judgement we were confronted with the broad nature of the survey question. Topics models reflected the data accurately but were not trained to look for the specific information that we were looking for. Therefore, distilling specific information about voluntary action, whether formal, informal, or potential proved to be somewhat challenging.

The model with 30 topics (as suggested by the perplexity measure) did end up being the most insightful to us. In this model, we were able to observe a topic (n=193) that seemed to be solely dedicated to more traditional interpretations of voluntary action. We have extracted all textual data assigned to that topic for closer inspection. Other topics contained a variety of interesting terms alluding to more ambiguous forms of volunteering and possible intentions, but these terms were spread out across all 30 topics. In this case, we did not consider manual inspection of all topics in hopes of finding what we were looking for; we would be spending a lot of time on topics and terms outside of our interests.

|  |  |  |  |
| --- | --- | --- | --- |
| **Term 1** | activ | **Term 6** | continu |
| **Term 2** | work | **Term 7** | support |
| **Term 3** | voluntari | **Term 8** | help |
| **Term 4** | local | **Term 9** | communiti |
| **Term 5** | involv | **Term 10** | church |

Table 3: Top 10 terms for topic 23

|  |  |
| --- | --- |
| **Text160** | I will still be working - probably still Full time - and still enjoying my work and Feeling it enriches my life. My children will now be 31 & 29 - I do not expect to be a grandparent yet but hope they are settled in relationships. We will still be very close - even if they live in other parts of the country/world we will maintain our relationship through emails/phone calls/visits etc. My mum will now be 87 & probably still living alone & active but needing much more support From myself & the Family generally. I would like to be more involved with some community or voluntary group - putting something back! I would only remarry with my childrens blessing as they mature into relationships themselves their tolerance & understanding will deepen. |
| **Text 233** | It's just my wife and I at home now. Our daughter has left after getting married. I have more time for tending the garden as I now work part-time as a clerical officer in a local firm based in {PLACE East of England} about two miles away. I need the extra income to top-up my pension. My wife and I are involved in the church that we've attended for some years now although I only preach once a year these days. My main role there is still a Homegroup leader caring and mentoring a small number of the other church members. My health is still good enough to carry out these aforementioned activities. I have again taken up "geriatric" badminton on a weekly basis. It's a full life! |
| **Text****305** | Retired, looking forward to taking out and caring for grandchildren. Reasonably healthy being able to walk for fair distances in the country and being interested in some form of hobby which is not relianment on physical activity as now. |
| **Text****332** | I IMAGINE I AM STILL DOING THE SAME JOB & AM HAPPILY MARRIED & STILL LIVING IN THIS AREA. MY HEALTH IS GOOD & I WILL STILL ENJOY SIMILAR ACTIVITIES & ENJOY A CLOSE RELATIONSHIP WITH MY CHILDREN, HOWEVER I WORRY ABOUT WHEN I WILL BE WORKING BEYOND 65. I WOULD MAYBE LIKE TO GET INVOLVED IN VOLUNTARY/COMMUNITY WORK IF TIME ALLOWS & WOULD LIKE TO INCREASE MY PHYSICAL ACTIVITIES. |
| **T348** | My physical health is excellent, regular gym, pilates & yoga sessions (several per week) have led to this. I am financially secure and able to help various friends & charities as a result. I remain interested in Financial Trading, and horse racing. Having succeeded in physical & financial areas, more focus now on spiritual practices & developments as I age. My loose network of friends continue to provide adequate social contact. |

Table 4: First 5 texts assigned to topic 23

*Considerations*

At this point, we were unable to identify specific topics of interest besides topic 23. However, the topic model did give us some insights into themes that might be hidden in the texts. As mentioned above, across topics we noticed possible references to both informal/invisible and potential voluntary action. Unfortunately for us, they were not grouped in a way that would represent dedicated topics. The noise was simply too high for interpretation.

Our choice for topic modelling was rooted in the bottom-up nature of the method. We considered this to be important as we had highlighted the need for a more contextual understanding. However, as our question was extremely broad, this meant a lot of noise (to us) was being brought into the model. If we were looking to answer the specific question that respondents were asked, the model would have likely been very successful for this task. In our case, we concluded we needed a more targeted approach.

When learning about computational text analysis, which was an ongoing process alongside the study itself, we came across another pitfall of this method for our purpose. LDA bases itself on the document term matrix that we had made before. When using a document term matrix, the assumption is made that order and syntax of words do not matter. This is called the ‘bag of words assumption’. Interestingly, models generally tend to perform rather well despite it being a false representation of language structures. Therefore, we had not considered this to be a concern before. However, now that we were evaluating our options, we started to consider another computational text analysis method where the context of word use could be preserved. In the next section, this other form of computational text analysis will be discussed and presented.

**5.2 Machine Learning Models**

Machine learning models (MLMs) are another branch of computational text analysis. Unlike topic models, which operate relatively unsupervised, MLMs are guided by human supervision. In this format, researchers know what they are looking for and communicate this to the model. The model then sets out to collect all datapoint of interest within the dataset.

In order to scan for relevant data, the researcher has to inform the model. This is done through a pre-coded sample. These codes are referred to as labels. Data features (in our case combinations of words) are considered predictors of these labels. The model is asked to infer the relationship between the input features and their related output class. To ensure the effectiveness of these models, training on held-out data takes place next for evaluation purposes. This is essentially the same process we went through with the perplexity measure, only now we were looking for the accurate prediction of our labels. General machine learning is rather similar to regular statistical modelling, where our input features (words) could be described as our independent variables. The output class (labels) are classifications of our dependent variable. However, MLMs are different from ‘normal’ statistical modelling in the sense that their sole aim is to predict, and interpretability is notoriously low. Models can have thousands of independent variables that do not make sense to the human eye yet are adding something to the prediction accuracy of the model.

There are multiple versions of such models in existence. We have started working with two relatively simple versions for text classification: Naive Bayes and Support Vector Machine. Studies have indicated that these two can be very effective tools, which require minimum complex knowledge on part of the researcher. As we were all new to this type of analysis, exploring those two is a good starting point. For the sake of reliability, consistency, and a comparison reference point to the topic model, we decided to follow guidelines from Van Atteveldt and Welbers (2022) again. We set out to follow their GitHub tutorial, which is publicly accessible at: https://github.com/ccs-amsterdam/r-course-material/blob/master/tutorials/machine\_learning.md

**5.2 MLMs and NCDS**

*Labelling*

Our initial dataset did not include any labels relevant to our study. Therefore, we had to manually code a data sample before we could start the process of model training. Although we had a slight sense of the contents of our data from the topic model, this was not extremely informative. Just making up codes before having a deeper understanding of the data would not be really logical. Moreover, by using ready-made top-down concepts, we would essentially ignore the rich bottom-up nature of our data set. Now that we had shifted to a supervised machine model, we had to be rather mindful of our own biases in order to build from the data and not vice versa. We approached carefully and with lots of deliberation.

A sample (n=200) was drawn from the data set and moved to AtlasTi, where manual coding was performed until saturation was achieved. Coding was structured according to a layered strategy proposed by Goia and colleagues (2013), which would allow for minimal researcher bias. This entailed a first step where first-order terms are coded in a detailed in vivo manner, preserving participants' texts. This step was followed by forming second-order themes, where researchers considered the established terms through a theoretical lens. In this final coding stage, these themes were grouped into three aggregate dimensions that encapsulated the central themes in the discourse that could be of interest to us. The labelling phase ended up taking much longer than expected due to questions surrounding the interpretation of concepts (aggregate dimensions). We have moved between several options before settling on our current labels (figure 2).



Figure 2: Example of the coding process

While labelling a total sample of 1400 texts, we ended up going back and forth on whether definitions would cover the totality of what we wanted to extract. We spend quite some time conceptualizing care: e.g., questioning if it is inherently a dedication of time and other resources, or if is it enough to wish others well. What does it mean to care? What about self-care? Similarly, there were concerns that the model would pick up more than we were interested in and data would be polluted. We wanted care to be a clear umbrella term that would bind all facets of it together in a mosaic fashion without blinding us to different colours. We settled on a broad interpretation of care encompassing references to traditional care work, references not just to caring actions but to the desire to help others, and wishing others well or worrying about their wellbeing. Our initial understanding of care also included enjoying spending time with others or actively engaging in group activities (e.g., sports club).

 Similar discussions followed on the care dimensions and what distinguished them from each other. E.g., could we assume that a wish for grandchildren meant that they also intended to care for said children, through babysitting, or did we need explicit mention of that? Throughout this process, we came to a general understanding of each dimension of care and set up guidelines that we could fall back on. E.g., in this question of grandchildren mentioned before we did want to have confirmation of some type of investment of time. An example of such guidelines is shared below (figure X) Even though we went on with a relatively fixed coding scheme, discussions on it have not stopped as interpretation remains a subjective matter.



Figure 3: Example of some coding guidelines

*Data Preparation*

Labelled data was loaded into R-Studio. Several data features were explored and adjusted for modelling purposes where needed. E.g.., textual data needed to be coded as factor class rather than character class. During preparation, it became apparent that there might have been a slight issue with either the sampling or the original data set as a few duplicates were detected. After removal, 1384 remained.

|  |  |  |  |
| --- | --- | --- | --- |
| **Care\_Family** | **Care\_Friends** | **Care\_Others** | **No label** |
| 902 | 314 | 203 | 393 |

Table 5: Label distribution in the training data

**5.1.1 Naive Bayes Model**

*Training*

The labelled data was then transformed into a corpus object and split into training and testing data, where about 80% of the data set was dedicated to training purposes. Further, a document term matrix was created, tokenizing and stemming our training data. Data was then trained.

*Testing*

Test data underwent the same process as the training data before: it was made into a corpus object and from there into a document term matrix.

*Evaluation Metrics*

For both models, we have calculated their respective metric values. Metrics are indicators that represent the performance of a model. Metrics are great tools to quantify the quality of output data and can be used to compare model performances. However, quality can be interpreted in a variety of ways. Therefore, we considered four types of metrics that all relate to the model's confusion matrix output.

A confusion matrix is a representation of the number of accurate predictions versus the number of predictions that were not accurate. In our matrices, zeros represent the absence of a label and ones represent the presence of a label.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Care\_Family* | *Care\_Friends* | *Care\_Others* |
|  | **Coded 0** | **Coded 1** | **Coded 0** | **Coded 1** | **Coded 0** | **Coded 1** |
| **Predicted as 0** | 68 | 30 | 161 | 20 | 175 | 13 |
| **Predicted as 1** | 30 | 152 | 44 | 55 | 60 | 32 |

Table 6: Confusion matrix NB model

Accuracy informs us how many times the model was right overall. It reflects on how many times a model incorrectly or correctly predicted a class. The precision metric considers what percentage of identifications turned out to be correct. The Recall metric reflects the proportion of true positives that were identified correctly. The F1 score can be seen as a harmonic mean between precision and recall scores.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Care\_Family** | **Care\_Friends** | **Care\_Others** |
| **Accuracy** | 0.79 | 0.77 | 0.74 |
| **Precision** | 0.69 | 0.89 | 0.93 |
| **Recall** | 0.69 | 0.79 | 0.74 |
| **F\_meas** | 0.69 | 0.83 | 0.83 |

Table 7: Metrics NB Model

**5.1.2 Support Vector Machine Model**

Testing and training were done on the same testing and training data. Metrics are represented in a similar matter. In this way performance between our two models could be compared.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Care\_Family** | **Care\_Friends** | **Care\_Others** |
| **Accuracy** | 0.78 | 0.90 | 0.87 |
| **Precision** | 0.71 | 0.92 | 0.91 |
| **Recall** | 0.70 | 0.95 | 0.93 |
| **F\_meas** | 0.71 | 0.93 | 0.92 |

Table 8: Metrics SVM Model

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Care\_Family* | *Care\_Friends* | *Care\_Others* |
|  | **Coded 0** | **Coded 1** | **Coded 0** | **Coded 1** | **Coded 0** | **Coded 1** |
| **Predicted as 0** | 69 | 28 | 194 | 17 | 219 | 21 |
| **Predicted as 1** | 29 | 154 | 11 | 58 | 16 | 24 |

Table 9: Confusion matrix SVM model

*Interpretation*

When considering the metrics of both models, they appear to function rather successfully. Especially accuracy is fairly high. Other measures appear slightly lower, yet high enough for functioning. However, when looking at the confusion matrices, a more complex picture is painted regarding these differences. For example, in table 6 above, the algorithm predicted it to be appropriate to label 30 cases that we ourselves considered irrelevant for the Family category. For that same category, 30 cases that we did label ourselves were dismissed by the model. These are our false positives and negatives. They are present for every category, which is essentially not an issue as this happens in most models. In the Family and Friends category, true positives and negatives are abundant. However, we did share some concerns over the predictions made in the Others category. The model seems better at predicting non-labels than labels (our dimensions). As a result, we could be missing out on a lot of relevant data points when we apply these models to the rest of our (unlabelled) data.

We proceeded to form a file of all the Care for Other cases, joined with their respective prediction and true labels. In this way, we were able to manually inspect the circumstances under which the models assigned certain labels to texts.

**6. Moving Forward/Future Work**
This initial phase of work does suggest that it is a promising technique for two reasons. Firstly it clearly represents an advance on frequency counts of terms chosen by the researchers (even though the patterns from some of those frequency counts are not without interest). Secondly it enables us to say something not just about the orientation of respondents towards voluntary action, but also about the relative priority they accord to it. Judging by the initial responses we do end up with groupings of responses which indicate that family and friends are prioritised over the dimension of care for others. This is not unlike the pattern of responses to survey questions described earlier in this paper but its advantage is that explore the content of individual responses in much more detail.

Further work is necessary and an obvious question to ask is whether those showing some interest in volunteering followed through on their intentions when aged 60. A further wave of the survey took place in 2013, so we can observe volunteering for the cohort at that point, and see whether there is a correspondence between the views people had aged 50 and their volunteering five years later. In fact the level of volunteering reported in that wave was almost identical to that reported in 2008. We haven’t, as yet, analysed the 2013 survey data to to see how the classification of responses in our work matches up to patterns of volunteering in 2013. For instance, did those in the “care\_other” category report unpaid voluntary work?

Unfortunately, plans to repeat the survey in 2018 (when the cohort became 60) were pushed back to 2020, and the survey was then modified to allow assessment of the impact of Covid. Therefore we don’t have a direct read-across to a wave of survey data captured in normal circumstances at the point at which individuals were ten years older. Using other survey data we can, however, say that volunteering levels for those aged 60 and over in the UK have remained more or less static, even when allowance is made for Covid-related perturbations (essentially there was a rise in informal volunteering because people were working from home and supporting co-residents in their communities, and conversely Covid restrictions meant that people could not volunteer in formal, organisational settings (other than online) for much of 2020 and parts of 2021).

So if volunteering levels are stable what is this analysis really telling us? We began the work thinking that the data would provide insights into attitudes to volunteering and how they translated into later behaviour. Given the lack of change of volunteering rates, there doesn’t seem to be an obvious connection between the lack of interest expressed by individuals in volunteering when they were 50, and the behaviour of the same group of people some years later. The position of volunteering in people’s priorities as something that comes after other commitments and activities is also unpromising. However the tentative evidence of gender divisions and an educational gradient could suggest some scope for action. Clearly it’s not going to be possible to raise volunteering levels unless we raise awareness of the need for volunteers, and perhaps the differences revealed by analyses of this kind can be helpful here.

Moving forward, what would further analysis look like? We would like to examine and compare the outputs of all models qualitatively. We are curious to see whether topic 23 from the topic model approach might have some overlap with categories from the machine learning approaches. We are eager to find out to what extent the machine learning models understood the assignment we gave them or if they perhaps picked up things we had not yet considered.

After this evaluation step, we might return to the drawing board for new label conceptualizations. When satisfied with conceptualization, but not with output, we shall dive deeper into the machine learning options. We are quite certain more advanced knowledge on the topic could lead us to fine-tune the model parameters to create a better ‘fishing net’.

Finally, when we are satisfied with the working of our model(s) of choice, we have the ambition of correlating its outputs with other survey data. For example, reported volunteering on a Likert scale (which is one of the earlier survey questions in the 2008 wave) could provide insights into the gap between qualitative and quantitative measurements.

7. Bibliography/References

C. (n.d.). *r-course-material/machine\_learning.md at master · ccs-amsterdam/r-course-material*. GitHub. <https://github.com/ccs-amsterdam/r-course-material/blob/master/tutorials/machine_learning.md>

*CLS | 1958 National Child Development Study*. (n.d.). <https://cls.ucl.ac.uk/cls-studies/1958-national-child-development-study/>

de Wit, A. (2020). ‘When I’m retired...’: Using topic modeling to analyze open-ended survey questions in the A Broader Mind Longitudinal Survey.

DPG Media Privacy Gate. (n.d.). <https://www.volkskrant.nl/nieuws-achtergrond/het-altruistisch-overschot~b1d7ff02/>

Elliott, J. (2012). Imagine you are 60. Centre for Longitudinal Studies. <https://cls.ucl.ac.uk/wp>content/uploads/2017/07/FINAL-CLS-WP-20126.pd

Elliott, J. (2022). The Craft of Using NVivo12 to Analyze Open-Ended Questions: An Approach to Mixed-Methods Analysis. *The Qualitative Report*, *27*(6), 1673-1687.

Hustinx, L., Grubb, A., Rameder, P. et al. (2022). Inequality in Volunteering: Building a New Research Front. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 33, 1–17.

Jacobi, C., Van Atteveldt, W., & Welbers, K. (2018). Quantitative analysis of large amounts of journalistic texts using topic modelling. In *Rethinking Research Methods in an Age of Digital Journalism* (pp. 89-106). Routledge.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, *78*, 15169-15211.

Kendall, J., Mohan, J., Brookes, N., & Yoon, Y. (2018). The English voluntary sector: How volunteering and policy climate perceptions matter. *Journal of Social Policy*, 47(4), 759-782.

Letwin, O. (2003) The neighbourly society: collected speeches 2001-3. London: Centre for Policy Studies. available at https://cps.org.uk/research/the-neighbourly-society-collected-speeches-2001-2003/

Musick, M.A., & Wilson, J. (2008) *Volunteers: A social profile*. Bloomington: Indiana University Press.

Pro Bono Economics (2023) Hands in the air like you just do care. Available from <https://www.probonoeconomics.com/hands-in-the-air-like-you-just-do-care>

Rooney, P. M., Steinberg, K., & Schervish, P. G. (2004). Methodology Is Destiny: The Effect of Survey Prompts on Reported Levels of Giving and Volunteering. *Nonprofit And Voluntary Sector Quarterly*, *33*(4), 628–654.

Santelli, F., Ragozini, G., & Musella, M. (2020). What Volunteers Do? A Textual Analysis of Voluntary Activities in the Italian Context. In D. F. Iezzi, D. Mayaffre, & M. Misuraca (Eds.), *Text Analytics* (pp. 265–276). Springer International Publishing. <https://doi.org/10.1007/978-3-030-52680-1_21>

Shachar, I.Y., Von Essen, J., & Hustinx, L. (2019). Opening Up the “Black Box” of “Volunteering”: On Hybridization and Purification in Volunteering Research and Promotion. *Administrative Theory & Praxis*, 41:3, 245-265.

Steinberg, K. S., Rooney, P. M., & Chin, W. (2016). Measurement of volunteering: A methodological study using Indiana as a test case. *Nonprofit and Voluntary Sector Quarterly*, *31*(4), 484-501.

Thatcher, M. (1981) Speech to Women's Royal Voluntary Service National Conference ("Facing the new Challenge") | Margaret Thatcher Foundation, available online at https://www.margaretthatcher.org/document/104551

Weber, M. (2021). How Do 50-Year-Olds Imagine Their Future: Social Class and Gender Disparities. *SAGE Open*, *11*(4), 215824402110615. <https://doi.org/10.1177/21582440211061567>

Wiepking, P. (2021). The Global Study of Philanthropic Behavior. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *32*(2), 194–203. <https://doi.org/10.1007/s11266-020-00279-6>