4th International Conference on Advanced Research Methods and Analytics (CARMA2022)
Universitat Politècnica de València, València, 2022
DOI: http://dx.doi.org/10.4995/CARMA2022.2022.15089

Implementing sentiment analysis to an open-ended questionnaire: Case study of digitalization in elderly care during COVID-19

Ida Toivanen [1], Venla Räsänen [1], Jari Lindroos [1], Tomi Oinas [1], Sakari Taipale[1 2]
[1]Faculty of Humanities and Social Sciences, University of Jyväskylä, Finland, [2] Faculty of Social Sciences, University of Ljubljana, Slovenia

Abstract

The rise of digital technology has enabled us to utilize even more integrated systems for social and health care, but these systems are often complex and time-consuming to learn for the end users without relevant training or experience. We aim to perform Named Entity Recognition based sentiment analysis using the answers of eldercare workers that have taken a survey about the effects of digitalization on their work. The collection of the panel survey data was carried out in two waves: in 2019 and 2021. For the sentiment analysis we compare these two waves to determine the effects of COVID-19 on the work of eldercare workers. The research questions we ask are the following: "Has technology affected eldercare workers' emotions in their work and how?" and "Has COVID-19 affected eldercare workers' views on digitalization in their work?". The main results suggest that criticism of modern technology persists through time – that is, before and after the pandemic the same type of negative and positive sentiments are manifested in the results. However, the familiarization with technology during COVID-19 seems to have been decreasing negative sentiments and increasing positive sentiments regarding digitalization. Due to the smallness of our data, more research should be conducted to make firmer conclusions on the matter.

Keywords: *eldercare work; digitalization; sentiment analysis; named entity recognition; BERT.*

1. Introduction

Health care and social services are usually considered to be one of the most important public services for the wellbeing of people. While assisting technologies aim to not only aid in accessing information but to form the basis for multi-tasking, the success of this process needs to be carefully examined (European Digital Agenda, 2022; Bartosiewicz et al., 2021). Although the ultimate aim for the vastly expanding digitalization is to offer better care for the patients and clients, the effects are often first experienced by the employees in social and health care services. Without proper research into the usability of the services that are intended for care professionals to operate, we may do more harm than good. Instead of decreased costs, improper incorporation of digitized tools may lead to increased costs and data security issues - but also exclusion may be prominent when trying to integrate social and health care systems in countries with low population density (Laitinen et al., 2018).

These drawbacks may be overcome more easily with carrying out user experience studies, using social and health care personnel as interviewees, that may supply more information about the difficulties of diverse types of employee groups (Ylönen et al., 2020). Especially older generations of eldercare workers find it often difficult to use information and communication technology (ICT) as a part of their work. Even more experienced and digitally skilled eldercare workers have expressed difficulties in engaging with their customers because of the growing need to report and use different technological systems. While attention has been given to this subject by several scholars (see Bartosiewicz et al., 2021; Laitinen et al., 2018; Ylönen et al., 2020; Seibert et al., 2020), the expanding need for digitalization of the social and health care sector has amplified during the COVID-19 pandemic and presents a need for continuing focus on the matter. While there are many roads to be explored, we have narrowed our approach to the elderly care sector. The research questions we ask are the following: "*Has technology affected eldercare workers' emotions in their work and how?*" and "*Has COVID-19 affected eldercare workers' views on digitalization in their work?*". Using two panel surveys, collected in two waves in 2019 and 2021, we implement a sentiment analysis to the answers to open-ended questions. Comparing the two waves we strive to make a distinction between sentiments concerning the use of technology at eldercare work before and after COVID-19.

Contemporary textual models designed for low-resource languages are still few and far between, which has prompted us to dig into the research of those languages. In this case we use Finnish language as an example to explore the difficulties (such as lack of data) of using low-resource languages in natural language processing tasks. We aim to utilize state-of-the-art language models (FinBERT) as the basis for model implementation and build up a novel use case for extended analysis.

2. Data

Our analysis is based on the first (2019) and second round (2021) of University of Jyväskylä survey on eldercare work, which is a new survey on the working conditions and digitalization of eldercare work collected by the Centre of Excellence in Research on Ageing and Care (www.jyu.fi/agecare) at the University of Jyväskylä. The aim of the survey was to collect information on the working conditions and use of ICTs (Information Communication Technologies) among eldercare workers in Finland. In this paper our focus is on textual answers for open survey questions, entailing 3971 data samples in total. We examine the answers to the open-ended question "*What kind of emotions related to the use of technology have been present in your work during the last week?*", which acts as the basis for tracking sentiments in the data. In addition to studying the feelings elicited using technology that the workers encounter in their every-day-life, it is important to conduct this type of analysis to attain information about working conditions, workload and the changes happening within their professional field.

3. Methods

In this study we refer to the analysis of emotional content by tagging sentiment vocabulary (by means of named entity recognition) as sentiment analysis. Named Entity Recognition (NER) task can be described as a sequence labeling (Virtanen et al., 2019), or token classification, task in which entities are tagged to study the way they are being used in the data. These entities can be, for example, locations, names, or dates (see Ruokolainen et al., 2020). This information extraction technique brings forth accentuated knowledge of the distinctive terminology groups there are in the data. To examine the first research question (*"Has technology affected eldercare workers' emotions in their work and how?"*), sentiment related glossary was NER tagged in the data. Example of sentiment NER tagging is visualized in Figure 1. For the second research question (*"Has COVID-19 affected eldercare workers' views on digitalization in their work?"*), we drew conclusions based on the frequencies of different NER sentiment tags and compare the frequencies between the datasets of 2019 and 2021 (Table 3). In this chapter we describe the data preprocessing, NER tagging process and model making relevant to this research.

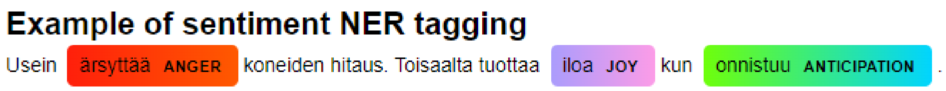

Figure 1. Example of sentiment NER tagging. The sentence translates to "Slow computers annoy (NER tag: ANGER) me often. On the other hand, it produces me joy (NER tag: JOY) when I succeed (NER tag: ANTICIPATION).".

3.1 Preprocessing and NER tagging

We preprocessed the data by removing empty and duplicate values. After preprocessing the dataset (of data from both 2019 and 2021) contains 3971 data samples. By data samples we mean one row of data that can consist of a single word, a sentence or multiple sentences. We first split the data to create a testing dataset of 10% (398 samples) of the entire dataset, to use it for the evaluation of the model. Then the remainder data was split into training (2858 samples) and validation (715 samples) datasets that were used during the model training process.

For NER tagging we used a Finnish sentiment corpus conducted by Mohammad, S. M., 2013. In the corpus there are eight sentiment classes (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) for which we built our word lists on. From the word lists irrelevant words were excluded, and appropriate synonyms were also added to the lists. We ended up with word lists of the following sizes: 212 words for anger, 123 for anticipation, 88 for disgust, 102 for fear, 134 for joy, 98 for sadness, 72 for surprise, and 31 for trust.

To start the comparing of sentiment word lists to our data, we first lemmatized, i.e., reduced a word to its basic word form, all the data samples. Each lemma in each lemmatized data sample was compared to each semantic word in every semantic word list (angry, sad, etc.), and when the word matched with a semantic word on the semantic word list, it was NER tagged and marked to be belonging to the semantic class (e.g., angry) at hand. This is how we made a new column of NER tags to prepare the data for token classification task. Entity type dataset statistics, or the amount of NER tags of different sentiment classes that are present in different datasets, are shown in Table 1. The whole data consists of approximately 8% sentiment NER tagged words.

Table 1. Entity type dataset statistics.

Entity	**Train**	**Valid**	**Test**	**Total**
Anger	690	213	90	993
Anticipation	597	118	63	778
Disgust	98	20	13	131
Fear	163	41	17	221
Joy	371	80	56	507
Sadness	189	47	22	258
Surprise	332	77	40	449
Trust	167	27	21	215
Total	2607	623	322	3552

3.2 Model

For building a language model that is supported by a low-resource language such as Finnish, we used the currently openly available state-of-the-art model, the FinBERT base (Virtanen et al., 2019) as the backbone and build a NER model upon that. We also used a ConvBERT (Jiang et al., 2020) model variation ConvBERT base Finnish, that is pretrained on a large Finnish corpus, as the backbone. Modest hyperparameter optimization was conducted while building the models. The models were trained for 10 epochs with a batch size of 16, learning rate of 5e-5 using a linear scheduler with two warmup steps, optimizer AdamW and weight decay set as 0, and maximum sequence length of 250.

4. Results

Results for NER task were obtained with several evaluation metrics: precision, recall, F1 score and accuracy. All values obtained with the evaluation metrics are shown in Table 2. The results suggest that ConvBERT model generally produces better results than the FinBERT model. This might be because replaced token detection (RTD) objective was used in the pretraining of the ConvBERT model (Jiang et al., 2020), while BERT's pretraining is based on the masked language modeling (MLM) objective (Virtanen et al., 2019).

Table 2. Results for testing dataset that is used to evaluate the model after the model training process is concluded.

Approach	Entity	Precision	Recall	F1	Accuracy
FinBERT	**Overall**	0.7892	**0.8424**	0.8150	0.9787
ConvBERT		**0.8208**	0.8392	**0.8299**	**0.9788**
	Anger	0.8068	0.8353	0.8208	
		0.8295	**0.8588**	**0.8439**	
	Anticipation	0.7727	**0.8500**	0.8095	
		0.8065	0.8333	**0.8197**	
	Disgust	1	0.7500	0.8571	
		1	0.7500	0.8571	
	Fear	0.7222	0.7647	0.7429	
		0.8125	0.7647	**0.7879**	
	Joy	0.9273	0.9107	0.9189	
		0.9811	**0.9286**	**0.9541**	
	Sadness	**0.8000**	**0.9091**	**0.8511**	
		0.6923	0.8182	0.7500	
	Surprise	0.6400	**0.8421**	**0.7273**	
		0.6522	0.7895	0.7143	
	Trust	0.7143	0.7143	0.7143	
		0.8889	**0.7619**	**0.8205**	

Additionally, we drew a subset of data from people who filled out the survey both in 2019 and 2021. We used this subset (n=1388) to compare how sentiment contents changed over the two-year period. The frequencies between 2019 and 2021 data proved to be so small that no comprehensive conclusions can be made from the results, shown in Table 3, alone. The biggest frequency decreases were in the sadness and anger entities, where the amount of NER tags were decreased by 22.22% and 18.57%, respectively. The amount of NER tags for trust and joy were conversely increased by 75.76% and 14.95%, respectively. This could implicate that the use of technology is more common among the eldercare workers after COVID-19, resulting in lesser use of sad and anger sentiments, that may be considered as negative sentiments, while reinforcing trust and joy sentiments, that can be seen as positive sentiments.

Table 3. Frequencies of words belonging to sentiment classes in a subset drawn from data from both 2019 and 2021 data.

Entity	2019	2021
Overall	713	709
Anger	210	171
Anticipation	142	157
Disgust	34	30
Fear	40	40
Joy	107	123
Sadness	54	42
Surprise	93	88
Trust	33	58

5. Discussion

The study at hand was carried out for two reasons: training language models to understand Finnish language and finding out the possible applications to different research settings. Prospects for the implementation of NER based sentiment analysis within the research field of social science (and others) are still largely to be discovered, especially because of the lack of trained tools for low-resource languages. Although the size of our dataset is relatively small, it still provides a basis to build domain-specific sentiment analysis. Identifying and describing difficulties that healthcare professionals face while using new technology at work is crucial to improve processes of software development and application in the future. Applying our approach could be used to rate the success of existing deployments – that is, to determine how accessible and usable the systems under evaluation really are.

Our preliminary results suggest ConvBERT model performs generally better than FinBERT model for our data. Additionally, we can make observations of the sentiment NER tag frequencies, that have changed from 2019 to 2021, that would suggest that COVID-19 has made eldercare workers familiarize themselves more with technology. More research with bigger data should be done to draw more precise conclusions. As our study is a retrospective analysis of these views, reflecting on them may provide useful results to demonstrate how more rigorous approaches to usability already ingrained in the research process are needed. For example, Jokela and Polvi (2010) present a way to conduct a test study by refining the usability requirements with as little ambiguity as possible. This was done by setting up a target level, a confidence value of 95 % that at least 75 % of the users will succeed in a task, that was supposed to be fulfilled or else they would have fallen short on the requirements. Jokela and Polvi (2010) make the conclusion that instead of asking about cosmetic details of the interface from test users, putting them under a test that is based on empiricism is needed for real user-friendliness. Defining requirements this way might make the process heavier to conduct but reduce the need for re-making the same systems in the future. Future research into these matters is required to form a comprehensive view on defining usability and how to serve eldercare workers with as little prejudice as possible.

The study includes also some limitations that should be taken into account. When designing a study to perform classification tasks, bias may be a fundamental part of the study when defining the classes. Especially in sentiment analysis, the way sentiments or single words in answers to open-ended questions are being divided into any of the eight classes already begs the question of the validity and reliability of our disposition to define the classes. As the finetuning of the tools is still in process with the more recent data, our latter part of the research is under thorough revision to meet up with scientific standards.

References

Bartosiewicz, A., Burzyńska, J., & Januszewicz, P. (2021). Polish Nurses' Attitude to e-Health Solutions and Self-Assessment of Their IT Competence. *Journal of clinical medicine*, 10(20), 4799.

European Digital Agenda (2022). *Digital Agenda for Europe*. Available online: https://www.europarl.europa.eu/ftu/pdf/en/FTU_2.4.3.pdf (accessed 21.3.2022).

Jiang, Z. H., Yu, W., Zhou, D., Chen, Y., Feng, J., & Yan, S. (2020). Convbert: Improving bert with span-based dynamic convolution. *Advances in Neural Information Processing Systems*, *33*, 12837-12848.

Jokela, T., & Polvi, J. (2010). Miten vaatia käytettävyyttä terveydenhuollon tietojärjestelmien tarjouspyynnöissä? Tapaus Oulun omahoitopalvelu. *Finnish Journal of eHealth and eWelfare*, 2(3), 129–135.

Laitinen, M., Hantunen, T., Heino, T., Hilama, P., Huttunen, A., Janhunen, P., . . . ammattikorkeakoulu, K. (2018). ”Digi vie, sote vikisee”: Kokemuksia sote-alan digitalisaatiosta DigiSote-hankkeessa Etelä-Savossa. Kaakkois-Suomen ammattikorkeakoulu Oy.

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word – emotion association lexicon. *Computational intelligence*, 29(3), 436-465.

Ruokolainen, T., Kauppinen, P., Silfverberg, M., & Lindén, K. (2020). A Finnish news corpus for named entity recognition. *Language Resources and Evaluation*, 54(1), 247-272.

Seibert, K., Domhoff, D., Huter, K., Krick, T., Rothgang, H., & Wolf-Ostermann, K. (2020). Application of digital technologies in nursing practice: results of a mixed methods study on nurses' experiences, needs and perspectives. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 158, 94-106.

Virtanen, A., Kanerva, J., Ilo, R., Luoma, J., Luotolahti, J., Salakoski, T., ... & Pyysalo, S. (2019). Multilingual is not enough: BERT for Finnish. *arXiv preprint arXiv*:1912.07076.

Ylönen, K., Salovaara, S., Kaipio, J., Tyllinen, M., Tynkkynen, E., Hautala, S., & Lääveri, T. (2020). Sosiaalialan asiakastietojärjestelmissä paljon parannettavaa: käyttäjäkokemukset 2019. *Finnish Journal of eHealth and eWelfare,* 12(1), 30–43.