Objective or subjective employment precariousness? Comparing definitions to a topic model based on user-generated data.

Ceciel Pauls [0000-0001-9452-1170], Michel Klein[2][0000-0003-4119-1846], Stef Bouwhuis[3][0000-0001-8554-360X] and Judith Bosmans[1] [0000-0002-1443-1026]

[1] Department of Health Sciences, Vrije Universiteit Amsterdam, Netherlands
[2] Department of Computer Science, Vrije Universiteit Amsterdam, Netherlands
[3]Department of Sociology, Vrije Universiteit Amsterdam, Netherlands
c.pauls@vu.nl

Keywords: Employment precariousness, Mental health, Topic modelling

1 Introduction

There exists great heterogeneity in the literature regarding both the definition and the operationalization of employment precariousness (EP) [1, 2]. Most authors agree that EP is a multi-dimensional concept. However, the literature on EP is defined by two distinct approaches to the concept [3]. On the one hand, EP can be defined as relating to the objective contractual arrangement, which we will refer to as 'objective EP' or relating to an individual's experience, which we will refer to as 'subjective EP' [3, 4]. In this abstract, we present our approach to use AI techniques, specifically classification algorithms and topic modelling, to produce a data-driven operationalization of EP.

The rise of social media to communicate has impacted the relationship between employees and employers [5]. However, limited research has been done to evaluate how employees might communicate about work on social media [6]. Social media data is an under-researched source of data to understand the ways in which work is experienced [7]. Thus, it is unclear how theoretically and data-driven definitions of EP, such as such EPRES-E and subjective EP, compare to data 'in the wild.'

In this study we aim to answer the following research question: How do prevalent topics discussed in relation to EP compare to both the objective and subjective definitions of EP? Additionally, we aim to determine whether and how we can determine the dimensions of both subjective and objective EP discussed in user-generated data.

2 Plan for data analysis

The proposed pipeline of our data analysis can be found in Figure 1.

2.1 Data collection and cleaning

We will develop a list of keywords based on definitions of EP used in scientific literature. This list of keywords will then be used to scrape posts and comments from work-related forums. Before performing any analyses, all documents will be tokenized. Additionally, the documents will all be formatted in lower case letters and emoticons, punctuation and numbers will be removed. Lastly, stop words (such as 'and' 'a' and 'the') will be removed. This will result in a dataset of documents (D).

2.2 Topic classification

We will use topic classification at multiple points in our study. First, we will use topic classification (Classification model 1, CM1) to identify relevant documents in which EP is discussed, thereby creating a dataset of relevant documents (D1). Second, we will use topic classification to determine the dimensions of both subjective EP (CM2) and objective EP (CM3) discussed in these relevant user-generated documents, resulting in two sets of labelled documents (D2 and D3). Each of our classification models (CM1 to CM3) will be trained on manually labelled datasets (L2 and L3). We will use ROBERTa (Robustly Optimized BERT Pretraining Approach) [8], a pretrained model of which the parameters can be fine-tuned by using labelled data [9]. In the past, ROBERTa has been used to classify Tweets regarding hybrid working as positive, neutral, or negative by Trivedi and Patel [10]. The performance of the models (CM1 – CM3) will be measured using Accuracy, Recall and F1-score.

2.3 Topic modelling

In this study, we will use a Bi-term topic modelling (BTM) approach to identify dimensions of EP that individuals use to discuss their employment online. By using the word-co-occurrence patterns in the whole corpus to generate topics, the issues with sparsity at document-level that occur when using methods such as LDA, are bypassed [11]. Bi-term topic modelling has been used to identify latent clusters of skill profiles in job listings [12] and thus might also be suitable for the current study. To determine the optimal number of topics, we will generate models ranging from 2 to 12 topics (K=2 to K=12). We will use both pointwise mutual information (PMI) and Diversity, to determine the quality of the model [13]. Additionally, we will focus on the interpretability of each topic, whether the topics identified are sufficiently distinct and the degree to which the words in the topics overlap [13]. After we have determined our final topic model (FTM), the dimensions identified by the model (FTM) will be compared to the dimensions of both objective and subjective EP.

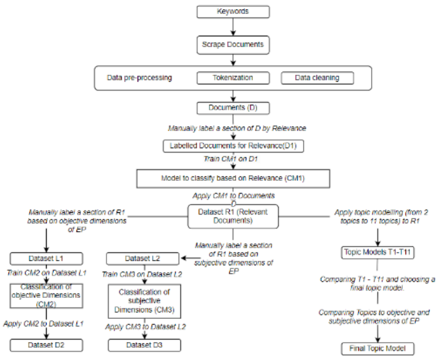

Fig. 1. Proposed pipeline

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this abstract.

References

1. Kreshpaj B, Orellana C, Burström B, Davis L, Hemmingsson T, Johansson G, Kjellberg K, Jonsson J, Wegman DH, Bodin T (2020) What is precarious employment? A systematic review of definitions and operationalizations from quantitative and qualitative studies. Scand J Work Environ Health 46:235–247
2. Jaydarifard S, Smith SS, Mann D, Rossa KR, Nikooharf Salehi E, Gnani Srinivasan A, Shekari Soleimanloo S (2023) Precarious employment and associated health and social consequences; a systematic review. Aust N Z J Public Health 47:100074
3. Kiersztyn A (2017) Non-standard employment and subjective insecurity: how can we capture job precarity using survey data? pp 91–122
4. Umney C, Kretsos L (2015) "That's the Experience." Work Occup 42:313–334
5. McDonald P, Thompson P (2016) Social Media(tion) and the Reshaping of Public/Private Boundaries in Employment Relations. International Journal of Management Reviews 18:69–84
6. van Zoonen W, Verhoeven JWM, Vliegenthart R (2016) How employees use Twitter to talk about work: A typology of work-related tweets. Comput Human Behav 55:329–339
7. Conway E, Rosati P, Monks K, Lynn T (2019) Voicing job satisfaction and dissatisfaction through Twitter: employees' use of cyberspace. New Technol Work Employ 34:139–156
8. Nozza D, Bianchi F, Hovy D (2020) What the [MASK]? Making Sense of Language-Specific BERT Models.
9. Devlin J, Chang M-W, Lee K, Toutanova K (2018) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
10. Trivedi S, Patel N (2022) Mining Public Opinion about Hybrid Working With RoBERTa. Empirical Quests for Management Essences 2:31–44
11. Yan X, Guo J, Lan Y, Cheng X (2013) A biterm topic model for short texts. In: Proceedings of the 22nd international conference on World Wide Web. ACM, New York, NY, USA, pp 1445–1456
12. Labussiere M, Bol T (2024) Are occupations "bundles of skills"? Identifying latent skill profiles in the labor market using topic modeling.
13. Churchill R, Singh L (2022) The Evolution of Topic Modeling. ACM Comput Surv 54:1–35