PACT-ML: Coding United Nation Peacekeeping Data from reports to the Secretary-General

Felix Kube*
Data Science Lab, Hertie School

May 14, 2025

Abstract

The Peacekeeping Activity Dataset (PACT) is the first of its kind data collection to shine light on what peacekeepers actually implement while deployed. In the past, many projects have looked towards mandates to study how specific tasks and mission success are related. PACT used report data from the mission heads to the Secretary-General of the UN to code up to 39 categories of task implementation on six different engagement levels. This project, PACT-ML, aims to extend the data collections of PACT 1.0 (Blair et al. (2022)) and PACT 2.0 (Otto (2024), Otto et al. (2024)) by using selected reports of PACT 2.0 to examine the application of Machine Learning / Natural Language Processing techniques to automatically code this sort of data from the reports.

Keywords: Machine Learning, Natural Language Processing, United Nations Peacekeeping, BERT, roBERTa

*Thanks for good advice during the semester, Chris and Killian. Also all the peers I talked to while tackling this group project alone :) it was not so alone after all.

1 Notes

- ONUCA reports left out due to super old report format and issues in preprocessing.
- maybe make cutoff at the introduction of the newer format?
- leave out cross-country missions due to problems in data parsing and different language format (only coded on state-level, when that state's name is explicitely mentioned)
  - could also preprocess data and merge those states back together

2 Introduction

Table 1: D-optimality values for design X under five different scenarios.

one	two	three	four	five
1.23	3.45	5.00	1.21	3.41
1.23	3.45	5.00	1.21	3.42
1.23	3.45	5.00	1.21	3.43

- Note that figures and tables (such as Table 1) should appear in the paper, not at the end or in separate files.

Motivation: Feasibility test to use data of this kind to automatically code newly written reports.

Background on the research project itself

3 Data

PACT 2.0 data, parsed reports

The UNPKOs included are: MINUGUA, MINUJUSTH, MINUSTAH, MIPONUH, ONUCA, ONUSAL, UNCPSG, UNCRO, UNMIBH, UNMIH, UNMIK, UNMISET, UNMIT, UNMOP, UNMOT, UNOMIG, UNPREDEP, UNPROFOR, UNSMIH, UNTAC, UNTAES, UNTAET and UNTMIH.

Include table of how many reports per mission in the whole dataset

Exclusion of very old reports due to issues in parsing

Exclusion of cross-country reports due to issues with the difference in language and adjacent codings, which we ignore for our language model approach.

Discuss how the data is structured (PACT 2.0)

Describe process how the data is parsed to extract the paragraphs

3.1 Validity checks

To be able to use our models on the data, we need to identify which paragraph text belongs to the specific coding in the PACT 2.0 data set. We take the reported paragraph, from which the manual coders at the University of Uppsala made their judgement, as ground truth. We do this because it is highly unlikely that mistakes happened, as each paragraph as per the UN reporting scheme carries its paragraph number at the beginning, and coders had to mark the relevant sentences within the PDFs before adding them to the database. On a side note, this fine-grained data which would allow us to train our models beyond the scope here, is only available for 143 out of 473 total reports. Three reports reported no

activity and can be therefore be left out for parsing.

To check if parsing was sucessful and our data quality is sufficient, we compare the number of paragraphs extracted from each PDF with the number of total paragraphs reported in the PACT 2.0 data, and only use the data for our model if the number of paragraphs align.

4 Methods

Don't take any of these section titles seriously. They're just for illustration.

5 Results

6 Conclusion

SUPPLEMENTARY MATERIAL

Title: Brief description. (file type)

References

Blair, R. A., Di Salvatore, J. & Smidt, H. M. (2022), 'When do UN peacekeeping operations implement their mandates?', *American Journal of Political Science* **66**(3), 664–680.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12650*

Otto, S. (2024), 'Peacekeeping Activity (PACT) Dataset 2.0'.
URL: *https://doi.org/10.7910/DVN/TQ8ETA*

Otto, S., Kube, F. & Smidt, H. (2024), 'UN peacekeeping upon deployment: Peacekeeping

activities in theory and practice', *Cooperation and Conflict* **59**(4), 488–509.

URL: *https://doi.org/10.1177/00108367241235888*