**Convokit Transformers**

Analysis

Testing

Training

Deployment

*Transformers-flavoured Unit Tests*

*Robustness across different works*

*Cornell Conv-Kit*

*Transformers Pipeline*

*Ablation Demonstration*

*Accelerate*

*Transformers Trainer*

Accessibility

*Transformers-flavoured Documentation*

*Cross framework support*

Figure 1: Outline of the capabilities of ConvoKit-Transformers

# **ConvoKit-Transformers: An applied study on Conversational Classification**

Project Proposal

Theophilus Lee\*

NUS High School of Math and Science

20 Clementi Ave 1, Singapore 129957

[h1910153@nushigh.edu.sg](mailto:h1910153@nushigh.edu.sg)

**Abstract**

*Studies on conversational classification have yielded nontrivial results in recent years. Significant milestones include the creation of DialogRE, conversation toolkits and increased adaptation of conversational knowledge. Such research would see a wide variety of applications to social media algorithms and psychology, among other fields. However, the current support for large scale application of such recent research is minimal at best. In addition, conversational classification has seen little adoption into widely used development toolkits. Implementation of research is typically for demonstration ability and not scalable for engineering purposes. We hence propose an applied study of recent methods used in the field, and compilation of work in a toolkit, ConvoKit-Transformers. We hold that our work will not only streamline research in the field, but also provide a practical evaluation of the current methods employed.*

# **Introduction**

Conversational classification is a subfield of Natural Language Processing that aims to extract classification labels from text in multi-speaker dialogue. This field has seen proposals for applications in social media algorithms and psychological evaluation, among many other current issues.

However, current resources for building models capable of conversation understanding lack many critical components for full stack software development. Clear documentation is not available for many proposed models, greatly inconveniencing a developer’s workflow. Little work exists to demonstrate the technical applications of conversational models to real world problems. We also find that many implementations of conversational models lack sufficient testing, and are highly self-contained, limiting robustness in a software development setting.[[1]](#footnote-1)

In this work, we intend to make conversational classification research more accessible for general application. We first reimplement TUCORE-GCN, a significant milestone of research in the field that has led to many derivative works proposed. We refactor the official implementation of TUCORE-GCN to include critical infrastructure necessary to engineering work. Next, we investigate the core necessities of a Machine Learning software developer when implementing a conversational model, delving into the needs of differing real-life scenarios and technical requirements among other aspects. Lastly, we package our work in a developer toolkit curated for conversational model creation, testing and deployment. We name this toolkit ConvoKit-Transformers, which extends the capabilities of ConvoKit, a previous conversation analysis toolkit. We release this toolkit to the public domain as an open-source project for research purposes.

Our main contributions can be summarized as follows:

1. We adapt the TUCORE-GCN branch of models for developer use.
2. We investigate the necessities of an ML developer for utilizing conversational models in solving conversational problems.
3. Our research is unified into an accessible and extendable toolkit, ConvoKit-Transformers

# **Previous Works**

We have seen recent progress in LLMs been applied to the field by performing retrievals from recent generative LLMs such as LLAMA2, reformatting emotional classification in conversation as a generative task. Tools to facilitate research in the field have been proposed. a dataset for Dialog Relation Extraction, a framework for formatting dialog relation inputs, a compilation of research implementations for emotional classification in conversation, and a toolkit for analysis of conversation datasets. Modelling of speaker information and turn attention approaches have shown improvements in classification tasks. Trigger Word detection was proposed as a possible heuristic for identifying emotions in speaker text. The field remains within theoretical grounds at large, since the lack of established resources hinders researchers from performing applied studies.

# **Proposed Methodology**

We take a multifaceted approach to outlining our research.

First, we adapt the TUCORE-GCN branch of models for developer use. TUCORE-GCN leveraged graph convolutional networks (GCN) for better understanding of entity-speaker-turn relations. A turn attention mechanism was also proposed. We consider TUCORE-GCN due to its relevance in the field and proven effectiveness.

To accomplish this, we reimplement TUCORE-GCN using the widely used HuggingFace transformers library. This awards developers the additional functionality offered by base classes of this library, and promotes modularity, allowing for easier reconfiguration of the model for downstream tasks. TUCORE-GCN consisted of implementations of the proposed architecture for BERT and RoBERTa. The resulting models were benchmarked etron accuracy, precision, f1, and f1c, a new metric proposed in dialogre paper which is the harmonic mean of conversational precision and recall. An ablation study was conducted on the model, removing each novel contribution and generating metrics for each produced model. However, we found that the initial implementation of the model had several issues. First, the data processing module. Input ids 2 and 3 were reserved for implementation of the special token used in BERTs. This has poor compatibility with widely used interfaces such as huggingface transformers, as the same tokens may be reserved for other purposes. Hence, we reworked the tokenization to allow for simple modification of the token ids to different values, allowing for greater robustness. We have additionally chosen more clear intermediaries, replacing the strings ‘unused2’ and ‘unused3’ with ‘entity\_1’ ‘entity2’, which is more in line with the representation demonstrated by the item.

-

We additionally redesign the data processing pipeline to function as an input processor for inference. We also propose a test suite in the transformers library style. This ensures that the work the code performs is accurate to research, and furthermore ensures sustanability of code. We document all functionality in the transformers library docstring formatting style, increasing ease of understanding for developers. To summarize the above improvements, we will produce a technical report on our implementation of TUCORE-GCN, covering any parity concerns and detailing how the work was achieved for future studies. We will additionally train TUCORE-GCN on models besides BERT and RoBERTa, and provide training details similarly to the TUCORE-GCN paper’s ablation studies. We will additionally publish code to generate such training details. Taking into light recent developments in research, we extend our implementation of TUCORE-GCN to Hi-Dialog, a derivative of TUCORE-GCN that replaces turn self-attention with hierarchical turn-based attention, among other improvements.

We then investigate the necessities of a Machine Learning developer when utilizing conversational models in solving conversational problems. We first perform an empirical study on current best practices in ML development. Specifically, we evaluate the design philosophy of current toolkits such as nltk or conv-emotion, and determine if they are applicable towards a toolkit for conversational classification. In addition. we investigate the needs of target users of our toolkit, and translate those needs into proposed functionality. Next, to demonstrate the capacity of conversational classification to solve real-world problems, we then prototype Machine Learning-based solutions to solve real world engineering problems. The scope of these prototypes is subject to change depending on the success of our TUCORE-GCN adaptation, although its main purpose is only to demonstrate applicability. These problems heavily differ in domain, to address a wide scope of issues that can be addressed by conversational classification.

Lastly, we package the results of our research into an accessible and extendable toolkit, ConvoKit-Transformers, taking into consideration the studies of other toolkits we have performed. The proposed toolkit will include generalized base functionality, allowing for extensions to other models, and will follow current standards in library design for ease of use. The toolkit will maintain parity between its model implementations and official implementations, or include an option to enable parity if it cannot be achieved. We build this toolkit around the base functionality of ConvoKit, a previously introduced toolkit for conversation analysis, that constructs a framework for conversation analysis. Convokit itself offers many datasets and utilities for analyzing conversational data, but is not well equipped for model creation, testing and deployment. We hence implement functionality to handle the additional workflow on top of Convokit’s abilities. The toolkit will be released on Github and PyPI, and will be free for non-commercial use.

# **Expected Outcomes**

The proposed toolkit will accelerate application of conversational classification models to real-world issues. Additionally, the toolkit will greatly facilitate research in the field, and increase understanding on conversational related fields.

# **Acknowledgements**

This research proposal is submitted in partial fulfillment of the NUS High School of Math and Science CS5131 Module Project Requirement. We extend our thanks to Mr Ng Chee Loong, for his guidance and support in this research. We also extend our thanks to the HuggingFace Community for their work in advancing open artificial intelligence research, and for heavily inspiring this research.

# **References**[[2]](#footnote-2)

1. Bongseok Lee and Yong Suk Choi. [Graph Based Network with Contextualized Representations of Turns in Dialogue](https://aclanthology.org/2021.emnlp-main.36). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 443–455, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics, 2021.
2. Xiao Liu, Jian Zhang, Heng Zhang, Fuzhao Xue, and Yang You. 2023. [Hierarchical Dialogue Understanding with Special Tokens and Turn-level Attention](https://arxiv.org/abs/2305.00262). In *ICLR*, 2023.
3. Jonathan P. Chang, Caleb Chiam, Liye Fu, Andrew Wang, Justine Zhang, and Cristian Danescu-Niculescu-Mizil. [ConvoKit: A Toolkit for the Analysis of Conversations](https://aclanthology.org/2020.sigdial-1.8). In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 57–60, 1st virtual meeting. Association for Computational Linguistics, 2020
4. Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. [COSMIC: COmmonSense knowledge for eMotion Identification in Conversations](https://aclanthology.org/2020.findings-emnlp.224). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2470–2481, Online. Association for Computational Linguistics, 2020
5. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](https://arxiv.org/abs/1810.04805). *arXiv preprint arXiv:1810.04805*, 2018.
6. Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. [RoBERTa: A Robustly Optimized BERT Pretraining Approach](https://arxiv.org/abs/1907.11692). *arXiv preprint arXiv:1907.11692*, 2019.
7. Po-Wei Lin, Shang-Yu Su, and Yun-Nung Chen. [TREND: Trigger-Enhanced Relation-Extraction Network for Dialogues](https://aclanthology.org/2022.sigdial-1.58). In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 623–629, Edinburgh, UK. Association for Computational Linguistics, 2022.
8. Weishi Wang, Steven C.H. Hoi, and Shafiq Joty. [Response Selection for Multi-Party Conversations with Dynamic Topic Tracking](https://aclanthology.org/2020.emnlp-main.533). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6581–6591, Online. Association for Computational Linguistics, 2020.
9. Dian Yu, Kai Sun, Claire Cardie, and Dong Yu. [Dialogue-Based Relation Extraction](https://aclanthology.org/2020.acl-main.444). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4927–4940, Online. Association for Computational Linguistics, 2020.
10. Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. [MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations](https://aclanthology.org/P19-1050). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 527–536, Florence, Italy. Association for Computational Linguistics, 2019.
11. Shanglin Lei, Guanting Dong, Xiaoping Wang, Keheng Wang, and Sirui Wang. InstructERC: Reforming Emotion Recognition in Conversation with a Retrieval Multi-task LLMs Framework. *arXiv preprint arXiv:2309.11911*, 2023
12. Ethan Zhou and Jinho D. Choi. [They Exist! Introducing Plural Mentions to Coreference Resolution and Entity Linking](https://aclanthology.org/C18-1003). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 24–34, Santa Fe, New Mexico, USA. Association for Computational Linguistics, 2018.
13. Sayyed M. Zahiri and Jinho D. Choi . Emotion Detection on TV Show Transcripts with Sequence-based Convolutional Neural Networks. In *Proceedings of the AAAI Workshop on Affective Content Analysis (AFFCON)*, 2018.
14. Tomasz Jurczyk and Jinho D. Choi. [Cross-genre Document Retrieval: Matching between Conversational and Formal Writings](https://aclanthology.org/W17-5407). In *Proceedings of the First Workshop on Building Linguistically Generalizable NLP Systems*, pages 48–53, Copenhagen, Denmark. Association for Computational Linguistics, 2017.
15. Kaixin Ma, Catherine Xiao, and Jinho D. Choi. [Text-based Speaker Identification on Multiparty Dialogues Using Multi-document Convolutional Neural Networks](https://aclanthology.org/P17-3009). In *Proceedings of ACL 2017, Student Research Workshop*, pages 49–55, Vancouver, Canada. Association for Computational Linguistics, 2017.
16. Henry Y. Chen, Ethan Zhou, and Jinho D. Choi. [Robust Coreference Resolution and Entity Linking on Dialogues: Character Identification on TV Show Transcripts](https://aclanthology.org/K17-1023). In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 216–225, Vancouver, Canada. Association for Computational Linguistics, 2017.

# **Appendix**

We further describe the current progress, and the resources and tools that will be used in conducting our research.

# **Project Plan and Timeline**

This research plan has been designed to be completed incrementally. The following tasks are self-contained, allowing for the project to be considered completed even if one or more features are not implemented. Sufficient documentation and testing is assumed to be compulsory for all tasks. Tasks where no obvious work is produced, such as conducting studies, will require producing a short report describing the work done.

* Reimplement TUCORE-GCN using the widely used HuggingFace transformers library *[Completed]*
* Redesign data processing pipeline to function as an input processor for inference. *[Completed]*
* Build a test suite in the transformers library style. *[Progress]*
* Document all functionality in the transformers library docstring formatting style *[Progress]*
* Produce a technical report on our implementation of TUCORE-GCN, covering any parity concerns and detailing how the work was achieved for future studies. *[Pending]*
* Produce TUCORE-GCN models trained on both BERT and RoBERTa. Consider producing models for edge devices. *[Pending]*
* Extend implementation of TUCORE-GCN to Hi-Dialog *[Pending]*
* Perform an empirical study on current best practices in ML development, applying those concepts to our work. *[Pending]*
* Investigate the needs of target users of our toolkit, and translate those needs into proposed functionality. *[Pending]*
* Prototype example solutions to solve real world engineering problems. *[Pending]*

The creation of the toolkit, Convo-Kit Transformers, is necessary for project completion.

# **Resources and Tools Used**

All models will be trained on a Nvidia RTX-3050 Laptop GPU, AMD Ryzen 7 5825U, and 4 GB GDDR6 RAM, unless otherwise specified. All model inference and training will be conducted in batches of 16, with gradient accumulation steps to test batches of 32 and 64 respectively. fp16 will be specified when used. quantization and low rank adaptation will be considered for producing models for edge devices. We will leverage the HuggingFace suite (transformers, accelerate, datasets, etc.) and Convo-Kit as the foundations of this research.

1. Equal contribution [↑](#footnote-ref-1)
2. <https://github.com/XiaoxinHe/Awesome-Graph-LLM> [↑](#footnote-ref-2)