# Hybrid and Generative Models for Material Science Event Extraction

**Anran Hao**[a], **Jian Su**[b]

[a] *Nanyang Technological University, Singapore 639798* S190003@ntu.edu.sg

[b] *Institute for Infocomm Research, A*STAR, Singapore 138632* sujian@i2r.a-star.edu.sg

## 1. Introduction

The vast amount of scientific knowledge stored in textual format, such as scientific papers, patents, and technical reports, creates a tremendous opportunity to develop and build information extraction tools to enable faster discovery, synthesis, and deployment into a wide variety of applications.

In this paper, we look into event extraction, a sophisticated type of information extraction, which is needed for knowledge discovery, but the state of art performances are far from satisfactory. In particular, we focus on event extraction from material science papers which lacks corresponding studies. We extend our hybrid deep learning model to material science event extraction, which equips syntactic information embedded in Large language models (LLMs) through its dependency tree-based residual connections for effective and efficient reinforcement. Our reinforcement units do not rely on external linguistic tools or domain-specific annotations, which are expensive to obtain. Experimental results demonstrate that our hybrid model achieves new state-of-the-art performance on the SC-CoMIcs benchmark dataset, establishing a high-performance baseline in the field.

We also evaluate the generative approach through a LLM model. While there remains a performance gap between the generative approach and our hybrid model, the work leads to the future explorations on the generative approach with even larger LLMs as well as the potential integration with the merits of both generative approaches and the hybrid deterministic model besides incorporating syntactic reinforcement into generative models. Besides, we also plan to extend our work to other material science event extraction tasks to support the experiment automation and knowledge discoveries in general, as well as other science disciplines, such as biology and climate.

## 2. Related Work

### *2.1 Material Science Information Extraction*

Despite the vast potential of NLP in material science, limited work has been done on event extraction in this domain. Early material science information extraction (IE) systems primarily focused on named entity recognition (NER) and relation extraction [1]. The first event extraction dataset for material science is SC-CoMIcs, which specifically focuses on doping events in superconductivity [2]. Our work is the first to introduce a syntactic tree-enhanced transformer model for event extraction in material science, setting a new high-performance baseline on SC-CoMIcs.

### *2.2 Event Extraction with Neural Architectures*

Event extraction (EE), which aims to detect and classify events and related argument mentions, is a crucial task in information extraction. It is also a challenging component in developing scientific informatics applications. Traditional approaches relied on rule-based methods and feature engineering [3, 4]. Recent advances in deep learning have significantly improved performances. SciERC is a widely used benchmark dataset for scientific entity and relation extraction [5]. For biomedical domain, state-of-the-art approaches such as [6] use discriminative deep learning models based on transformers. Event extraction, which requires capturing complex relationships beyond entity-relation pairs, has remained underexplored in material science.

While syntactic information is important for event extraction, incorporating it has been a longstanding challenge in EE. Previous works have explored various methods, including Graph Convolutional Networks [7] and Graph Attention Networks [8] for syntax-aware NLP tasks. In biomedical EE, tree-based syntactic structures have been used to improve performance [9]. We extend our hybrid reinforcement mechanism to EE from material science text, integrating syntactic dependencies into LLMs, without relying on external parsers while ensuring effectiveness and efficiency [10].

### *2.3 Large Language Models for Scientific NLP*

Pre-trained large language models (PLMs) such as BERT [11], and RoBERTa [12], and domain-specific models like SciBERT [13] and MatBERT [14] have demonstrated effectiveness in scientific text understanding. Recently, generative models such as GPT-4 [15] and instruction-tuned models like InstructGPT [16] have been explored for knowledge extraction tasks. While generative approaches offer flexibility, they often lack the structured precision required for event extraction [17]. Our study compares the hybrid approach with generative models, showing that while LLMs alone do not yet surpass structured extraction methods, integrating them with syntactic reinforcement may potentially improve performance further besides model level integration as well as the evaluation of larger LLMs.

## 3. Task Formulation

Following Yamaguchi et al. [2], we formally define the task as: Given input text S and a set of predefined trigger and argument types, which consists of a set of words $w = \{w_1, w_2, ..., w_n\}$, the model extracts all the event mentions by detecting spans $\{w_i\}$ that are event triggers and arguments and classifying them into respective types $\{y_i\}$. For example, in the sentence "Pb doping at the Bi site in $Bi_2Sr_2CaCu_2O_8$ increases $T_n$ (a key superconducting property of interest).", the output is expected to be a Doping event in which trigger-Doping="doping", argument-Dopant="Pb", argument-Site="Bi".

## 4. Our Methods

We extend our work [10] and adopt the PLM-model architecture with Soft Syntactic Reinforcement (SSR) mechanism. As shown in Fig. 1, we encode abstracts in segment with a PLM. As shown in Fig. 1 and 2, we use a dependency dataset to train an SSR component that learns to reconstruct dependency parse trees, and then use the syntax-sensitive transformation matrices to build a modified residual connection before feeding to a conditional random field (CRF)-based classifier.

In addition, we test out a generative EE model from the latest benchmark evaluation [18].

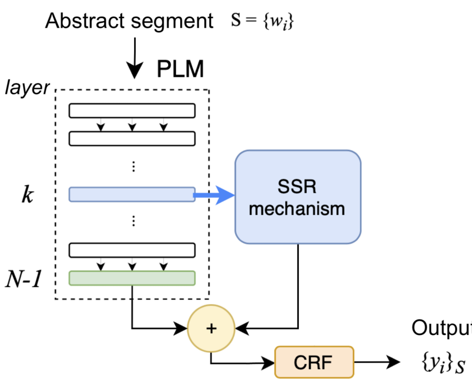


Fig. 1: Overall architecture of our proposed model.

## 5. Experiments

### 5.1 Experimental Setup

**Dataset.** We use the SC-CoMIcs dataset [2], which collects 1000 ScienceDirect research abstracts related to superconductivity and provides 2778 doping event annotations.[1]

**Baselines.** We consider the DYGIE++ model [19], which is the baseline used in the dataset paper [2], and the state-of-the-art discriminative and generative event extraction models from the latest benchmark evaluation [18], which are denoted as D-LM and G-LM in Table 1, respectively.

[1]The original dataset paper reported joint information extraction performance, but since NE and relation annotations are expensive and not always available for material science applications, we use the single task setting for event extraction evaluation.

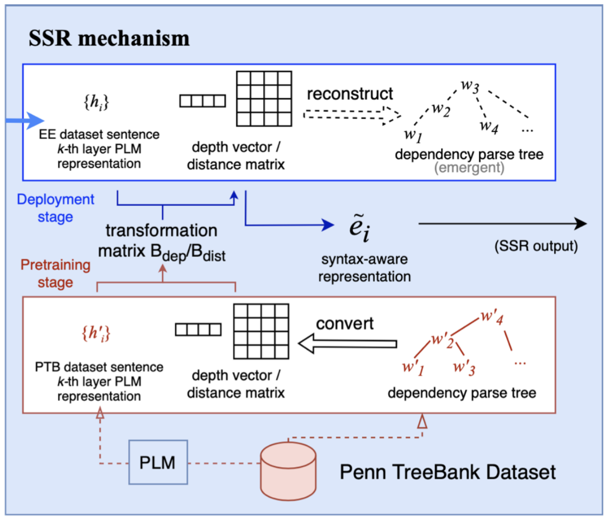


Fig. 2: Detailed illustration of our proposed Soft Syntactic Reinforcement (SSR) mechanism.

**Evaluation metrics.** We use precision (P), recall (R), and F1 as evaluation metrics. For other details, we follow the benchmark evaluation [18].

### 5.2 Experimental Results

Table 1 shows that our SRE model outperforms all other models compared. We also observe that there is a significant performance gap between the generative approach (G-LM) and the discriminative approach (Baseline, D-LM and SRE).

Table 1: Performance results on the SC-CoMICs dataset. The best results are highlighted in bold.

| Tasks | Model | P | R | F1 |
|---|---|---|---|---|
| Trigger Extraction | Baseline | 86.9 | 98.4 | 92.3 |
| | D-LM | 92.0 | **99.5** | 95.6 |
| | G-LM | 86.6 | 90.9 | 88.7 |
| | SRE (ours) | **95.2** | 98.4 | **96.8** |
| Argument Extraction | Baseline | 79.2 | 54.1 | 64.3 |
| | D-LM | 88.7 | 76.1 | 81.9 |
| | G-LM | 38.6 | 43.2 | 40.8 |
| | SRE (ours) | **89.0** | **81.3** | **84.9** |

## 6. Future Directions in Hybrid Approaches

Hybrid models that combine structured extraction with LLMs have been gaining attention. Recent work has explored methods integrating rule-based methods with GPT-based approach for legal information extraction [20]. In future work, we plan to explore hybrid frameworks that combine the strengths of generative models with syntactic reinforcement, aiming for further improvements in material science EE besides the incorporation of larger LLMs.

## Acknowledgments

This research is partially supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme for the programme DesCartes.

## References

[1] Yu Song, Santiago Miret, and Bang Liu. MatSci-NLP: Evaluating scientific language models on materials science language tasks using text-to-schema modeling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2023.

[2] Kyosuke Yamaguchi, Ryoji Asahi, and Yutaka Sasaki. SC-CoMIcs: A superconductivity corpus for materials informatics. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2020.

[3] Yi Luan et al. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018.

[4] Shiyao Cui et al. Edge-enhanced graph convolution networks for event detection with syntactic relation. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020.

[5] Zhijiang Guo, Yan Zhang, and Wei Lu. Attention guided graph convolutional networks for relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019.

[6] Anran Hao, Haohan Yuan, Siu Cheung Hui, and Jian Su. Effective type label-based synergistic representation learning for biomedical event trigger detection. *BMC Bioinformatics*, 25(1):251, 2024.

[7] Diya Li and Heng Ji. Syntax-aware multi-task graph convolutional networks for biomedical relation extraction. In *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, 2019.

[8] Lianzhe Huang et al. Syntax-aware graph attention network for aspect-level sentiment classification. In *Proceedings of the 28th International Conference on Computational Linguistics*, 2020.

[9] Hao Fei, Yafeng Ren, and Donghong Ji. A tree-based neural network model for biomedical event trigger detection. *Information Sciences*, 512, 2020.

[10] Anran Hao, Jian Su, Shuo Sun, and Yong Sen Teo. Soft syntactic reinforcement for neural event extraction. In *The 2025 Annual Conference of the Nations of the Americas Chapter of the ACL (Main Conference)*, 2025.

[11] Jacob Devlin et al. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019.

[12] Yinhan Liu et al. Roberta: A robustly optimized bert pretraining approach, 2019. cite arxiv:1907.11692.

[13] Iz Beltagy, Kyle Lo, and Arman Cohan. SciBERT: A pretrained language model for scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019.

[14] Nicholas Walker et al. The impact of domain-specific pre-training on named entity recognition tasks in materials science. *SSRN Electronic Journal*, 2021.

[15] OpenAI. Gpt-4 technical report. 2023.

[16] Ouyang Long et al. Training language models to follow instructions with human feedback. *ArXiv*, abs/2203.02155, 2022.

[17] Derong Xu et al. Large language models for generative information extraction: A survey. *Frontiers Comput. Sci.*, 18:186357, 2023.

[18] Kuan-Hao Huang et al. TextEE: Benchmark, reevaluation, reflections, and future challenges in event extraction. In *Findings of the Association for Computational Linguistics: ACL 2024*, 2024.

[19] David Wadden et al. Entity, relation, and event extraction with contextualized span representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019.

[20] Harry Nan, Maarten Marx, and Johan Wolswinkel. Combining rule-based and machine learning methods for efficient information extraction from enforcement decisions. In *International Conference on Legal Knowledge and Information Systems*, 2024.

## Appendix A. Base Language Models of the Compared Models

In our experiments, our Syntactic Reinforced model for Event extraction (SRE), baseline DYGIE++ and D-LM are finetuned based on RoBERTa-large, while G-LM is finetuned based on BART-large.