NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

School of Computer Science and Engineering
College of Engineering

Event Detection for Biomedical Text

Student: Pham Nguyen Minh Thu

Supervisor: Assoc Prof Hui Siu Cheung

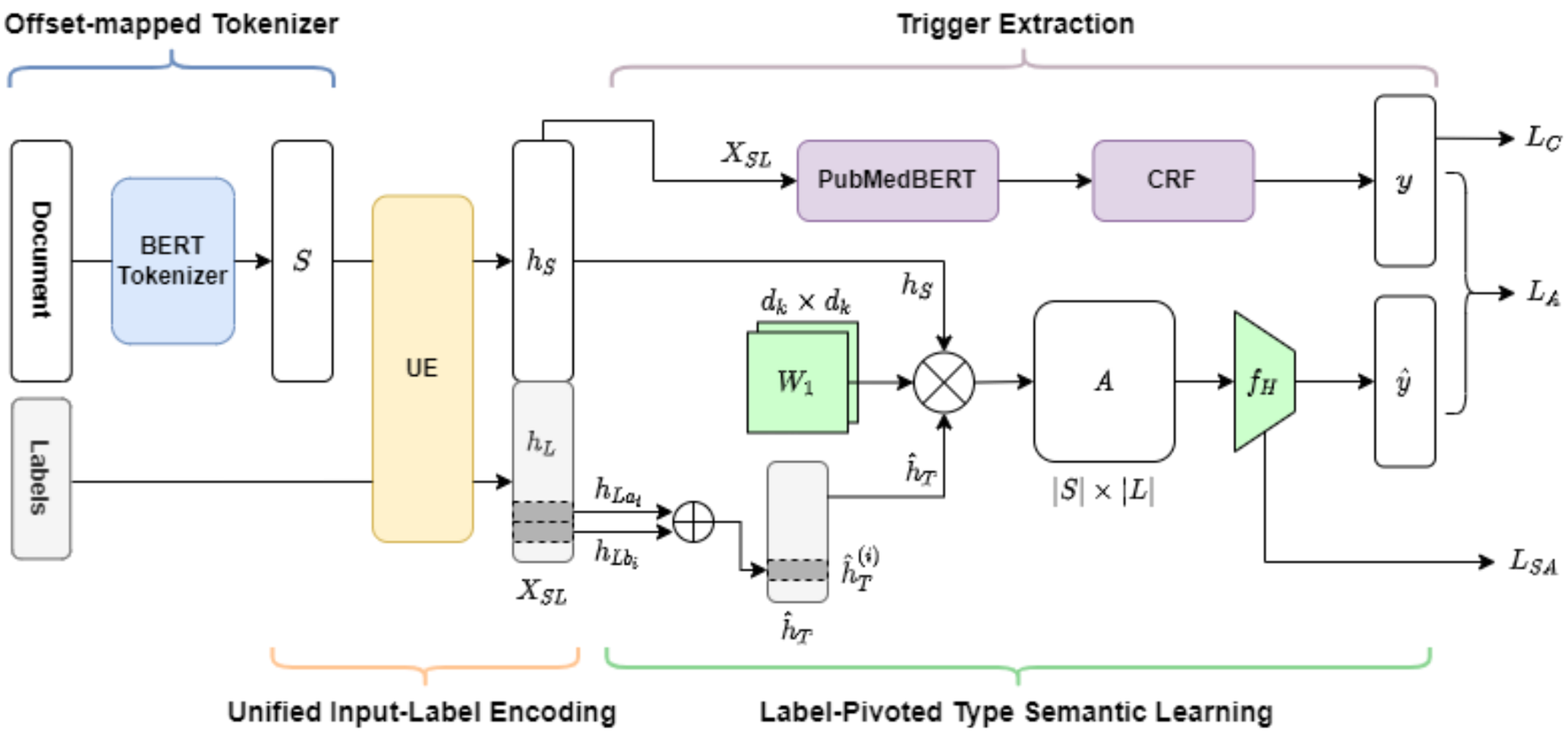

Figure 1. Architecture of our proposed LPBED model.

Objectives:

This research aims to apply Natural Language Processing to automatically detect event triggers in biomedical domain. Event detection is the primary step in the event extraction task, whose objective is to detect events via trigger mentions that best indicate the occurrence of events with a particular type. In this research, we propose a novel Label-Pivoting Biomedical Event Detection model (LPBED) which is pretrained with PubMedBERT language model and exploits the semantic meaning of the type label set. More specifically, inspired by the Semantic Pivoting Model for Effective Event Detection (SPEED) proposed by a PhD student Hao Anran, our proposed model makes use of the underlying semantic meaning of type labels to pivot event types as clues for detecting trigger candidates. Our model gains significant benefits from the pretrained PubMedBERT model for the domain-specific knowledge of the widely-used biomedical data sources. The proposed model serves as the groundwork for an extension to event extraction in biomedical domain.

Model Architecture:

Our proposed model consists of four modules:

- Offset-mapped Tokenizer: tokenizes the document into a token list attached with offsets mapping, which facilitates the conversion to the standard MAVEN format.
- Unified Input-Label Encoding: encodes the input sentence and the type labels into a joint representation, enabling interaction between the input tokens and the semantic clues of type labels through attention.
- Label-Pivoted Type Semantic Learning: implements a contrastive loss function to segregate negative mentions from the groups of matched candidates pivoted by its labels.
- Trigger Extraction: produces the final type prediction for each token based on sequence tagging for the input sentences.

Performance Results:

Model	**F1** (%)	**P** (%)	**R** (%)
BERT-CRF	68.03	71.11	65.22
LPBED	**68.90**	**71.48**	**66.51**

Table 1. Performance on dev set of GE11 dataset.