NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Computer Science and Engineering
College of Engineering

Event Detection for Cybersecurity News Articles

Student: Jovan Huang Tian Chun

Supervisor: Assoc Prof Hui Siu Cheung

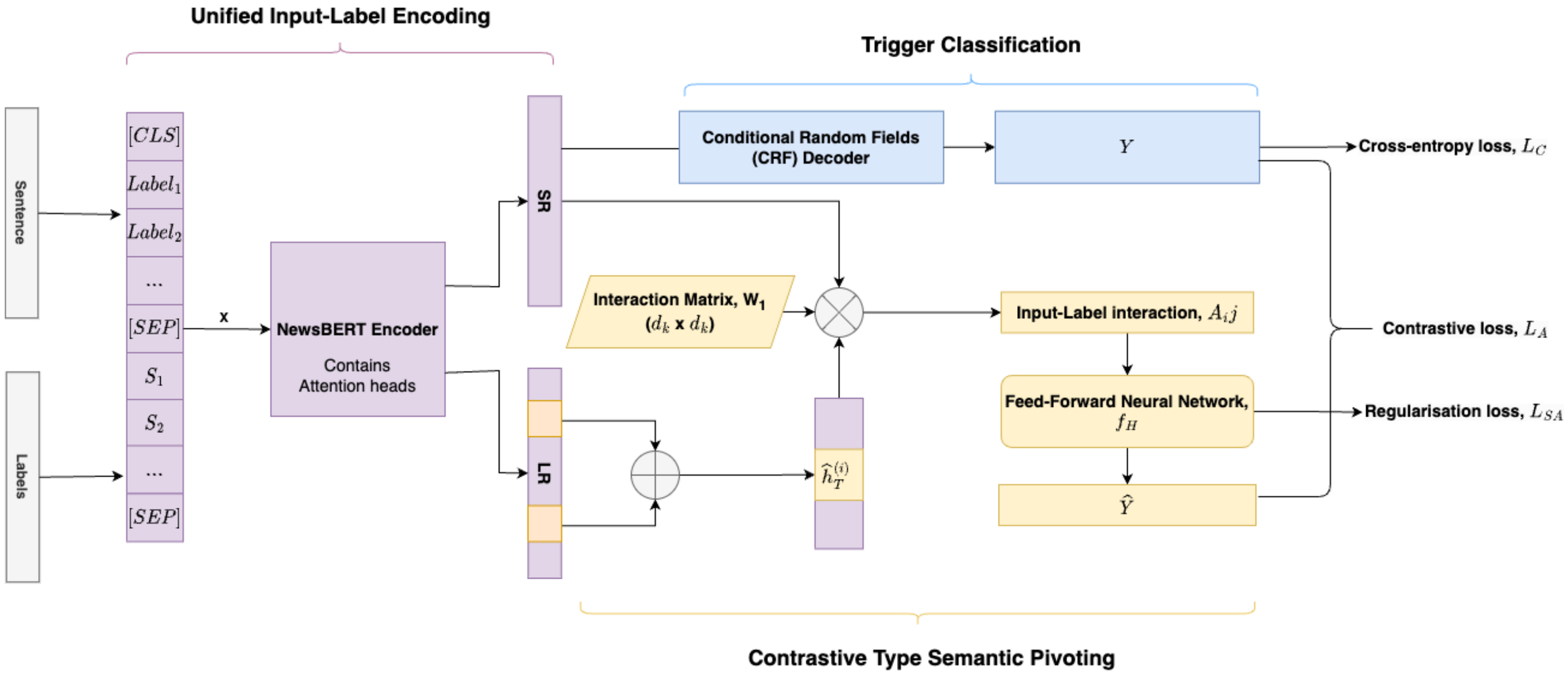

Figure 1. Architecture of our proposed LPCNED model.

Project Objectives:

This research aims to automate the detection of event triggers in cybersecurity news article domain utilizing Natural Language Processing techniques. The event detection process involves identifying events through trigger mentions that are most indicative of the occurrence of events belonging to a specific category.

The research proposes a new approach called Label-Pivoting Cybersecurity News Event Detection (LPCNED) that enhances the detection of cybersecurity events in news articles by incorporating semantic type representations. LPCNED builds on the Semantic Pivoting Model for Effective Event Detection (SPEED) proposed by Phd student, Hao Anran, but replaces the BERT-base encoder with a pre-trained NewsBERT encoder in the Unified Input-Label Encoding component to generate more effective input and label representations. The aim is to leverage the benefits of pre-training and utilize a task-specific architecture to improve the performance of SPEED for better event detection in cybersecurity news articles.

Model Architecture:

Our proposed model consists of three modules:

- **Unified Input-label Encoding:** Utilizes a NewsBERT contextualized encoder to generate input and label representations in a unified manner. The encoder facilitates the interaction between the input tokens and semantic clues of type labels via attention mechanisms.
- **Contrastive Type Semantic Pivoting:** Leverages a contrastive loss function to enhance the interaction between inputs and labels and produce type semantic representations.
- **Trigger Classification:** Employs a Conditional Random Field (CRF)-based decoder to generate predicted labels for input tokens.

Performance Results:

Model	F1 (%)	Precision (%)	Recall (%)
BERT-CRF	75.36	74.63	76.10
SPEED	75.44	73.40	**77.60**
LPCNED	**75.72**	**76.78**	74.69

Table 1. Performance on test set of CEAC dataset.