

## Trustworthy Disaster Summarization

Given large volumes of crisis-related tweets posted during the unfolding of sudden adversities, the task is to generate *trustworthy* and actionable summaries.



### Paris Terror Attack (October 2020)

- T1:** The Muslim who beheaded a teacher in a street in France waited outside the school and asked pupils to identify his target, anti-terrorism officials say. – **0.91**
- T2:** In an horrific attack outside Paris a teacher having taught a class about 'Freedom of Expression' beheaded by an 18yr old student, a Chechen, born in Moscow who filmed his attack before being shot dead by police at the scene! 9 people arrested include his family members! <link> – **0.86**
- T3:** The teacher brutally murdered for doing his job in France yesterday has been named as 47-year-old Samuel Paty. <link> – **0.81**

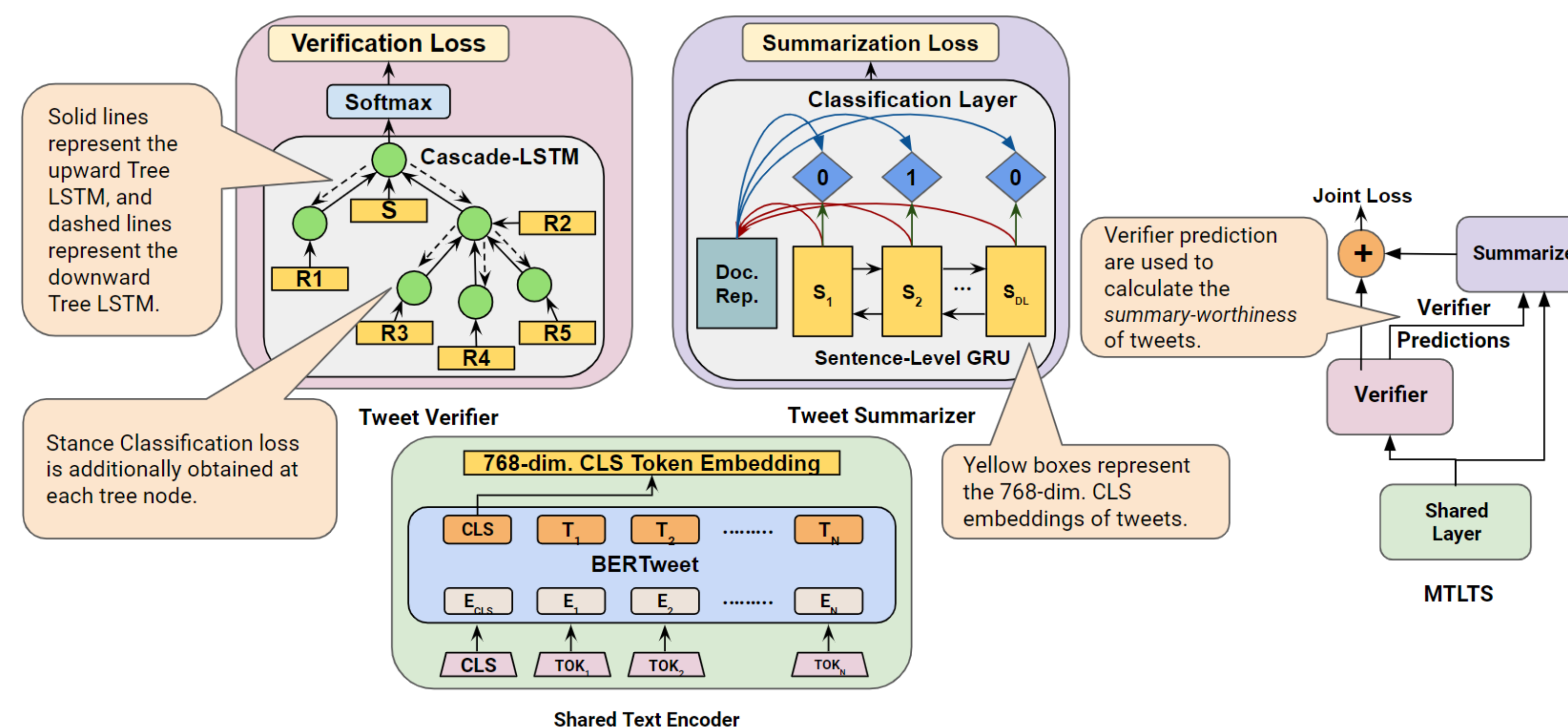
Joint score representing the authenticity and summary-worthiness of tweets.

### Our Goals

- To design an end-to-end solution for **trustworthy summarization** of disaster-related tweets.
- To take a **supervised** approach as against existing disaster-specific summarization algorithms.
- To address the gap between a vast body of literature on document summarization and comparatively limited research on social media summarization.

## MTLTS Architecture

- Tweet Summarization** is the main task. **Tweet Credibility Verification** becomes a very important auxiliary task.
- Our proposed solution **MTLTS** is trained using a hierarchical multi-task approach. The **verifier** is trained at a lower layer and its predictions are used to train the **summarizer** at a deeper layer for obtaining the summary-worthiness of tweets.



## Summary Evaluation and Comparison

Model	Charliehebd0		Germanwings		Ottawa		Sydney		Overall	
	V-Ratio	R1-F1	V-Ratio	R1-F1	V-Ratio	R1-F1	V-Ratio	R1-F1	V-Ratio	R1-F1
APSAL	0.200	<b>0.493</b>	0.056	0.437	0.000	0.435	0.000	0.483	0.064	0.462
COWTS	0.385	0.479	0.111	0.504	0.118	0.427	0.087	0.479	0.175	0.472
SCC	0.365	0.465	0.244	0.505	0.212	0.444	0.308	0.492	0.282	0.477
VERISUMM	0.722	0.381	0.624	0.519	0.645	0.396	0.619	0.378	0.653	0.419
MTLTS ( $\kappa = 1$ )	0.750	0.444	0.927	0.565	<b>0.860</b>	<b>0.477</b>	0.627	<b>0.518</b>	0.791	<b>0.501</b>
MTLTS ( $\kappa = 0$ )	<b>0.947</b>	0.472	<b>1.000</b>	<b>0.575</b>	0.834	0.464	<b>0.750</b>	0.452	<b>0.883</b>	0.491

Figure: Quantitative comparison of summaries generated by various methods on the *PHEME* dataset. *V-Ratio*: Verified Ratio of summary tweets.

### Baselines:

- APSAL**: A clustering-based multi-document summarization system for disaster-related articles.
- COWTS**: An extractive summarization approach to generate situational summaries from disaster-related tweets.
- SCC**: An extractive summarization approach to identify sub-events from disaster-related tweets, and summarize them.
- VERISUMM**: A 2-stage pipeline approach to verify and summarize disaster-related tweets.

## Summary Membership and Generation

$$p_i^{D_k} = P(y_i^k = 1 | h_i^k, \text{sum}_i^k, d_k, \text{Ver}_i^{\text{prob}})$$

$$\text{Summ}_i^{\text{prob}} = \frac{1}{m} \sum_{j=1}^m (p_i^{D_j} | s_i \in D_j)$$

$$\text{Score}(s_i) = \kappa \cdot \text{Summ}_i^{\text{prob}} + (1 - \kappa) \cdot \text{Ver}_i^{\text{prob}}$$

- $p_i^{D_k}$  represents the local probability of source tweet  $s_i$  being included in the summary of document  $D_k$ .
- $m$  represents the no. of documents containing  $s_i$ .
- The **relevance-score** of each source tweet  $s_i$  is calculated by combining its confidence/membership scores generated by the **verifier** and the **summarizer**.

## Key Highlights

- MTLTS is the first end-to-end solution to generate **trustworthy** and **actionable** summaries from potentially rumourous crisis-related tweets.
- Different from existing disaster-specific summarizers, our approach is **supervised**. Also, our proposed model is **generic** and can be applied to other domains.
- We present a novel method to leverage a document summarizer for summarizing social media posts.
- We achieve **state-of-the-art** results both for the primary task of trustworthy summarization as well as the auxiliary task of tweet credibility verification.
- We discover that the **veracity** of different classes of tweets **change over time** in different proportions as official details are gradually made public.



Contact Information



Paper Details