A Weakly Supervised Propagation Model for Rumor Verification and Stance Detection with Multiple Instance Learning

Ruichao Yang
Hong Kong Baptist University
Hong Kong SAR, China
csrcyang@comp.hkbu.edu.hk

Jing Ma
Hong Kong Baptist University
Hong Kong SAR, China
majing@comp.hkbu.edu.hk

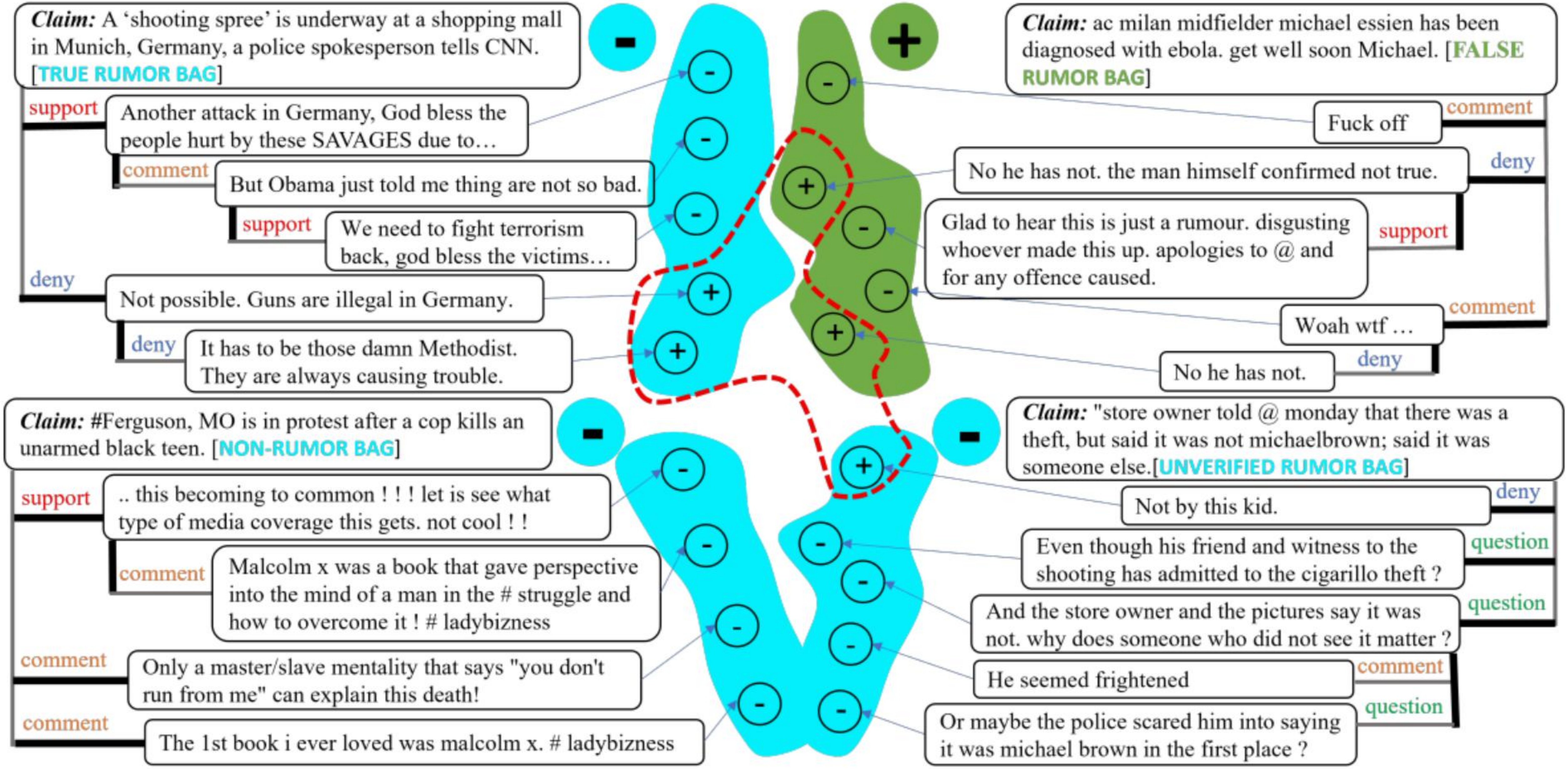
Hongzhan Lin
Beijing University of Posts and Telecommunications
Beijing, China
linhongzhan@bupt.edu.cn

Wei Gao
Singapore Management University
Singapore
weigao@smu.edu.sg

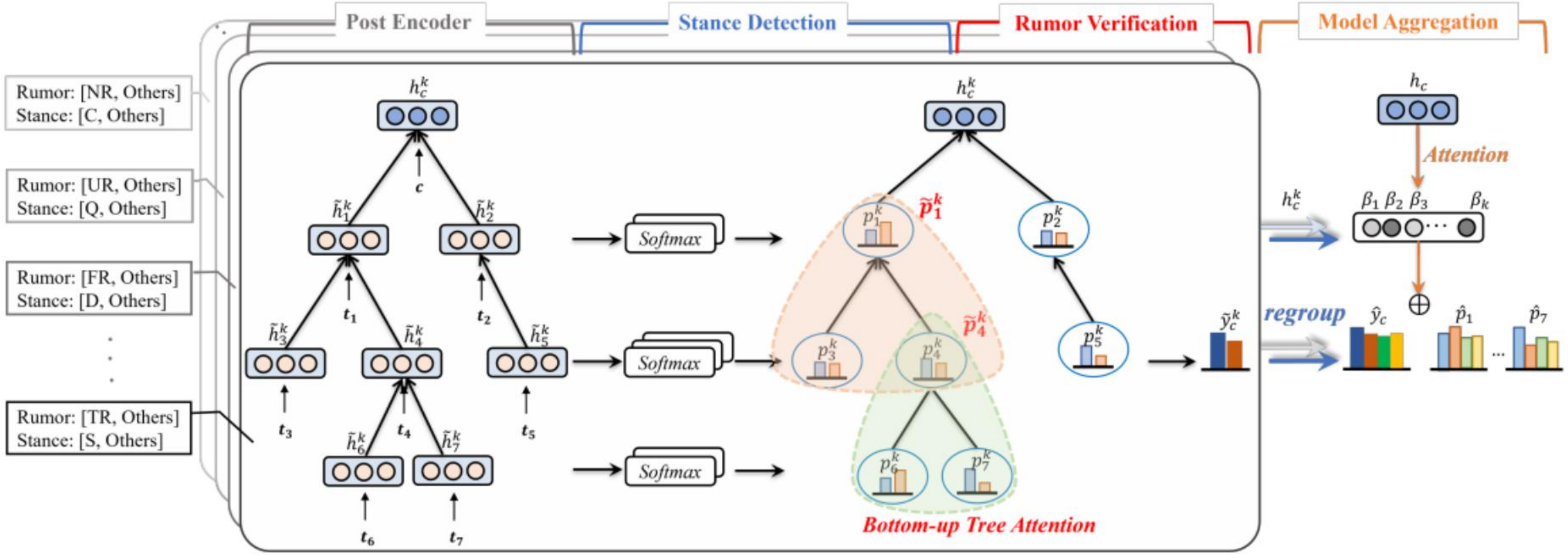
SIGIR2022

Reported by Xiaoke Li

Introduction



Method

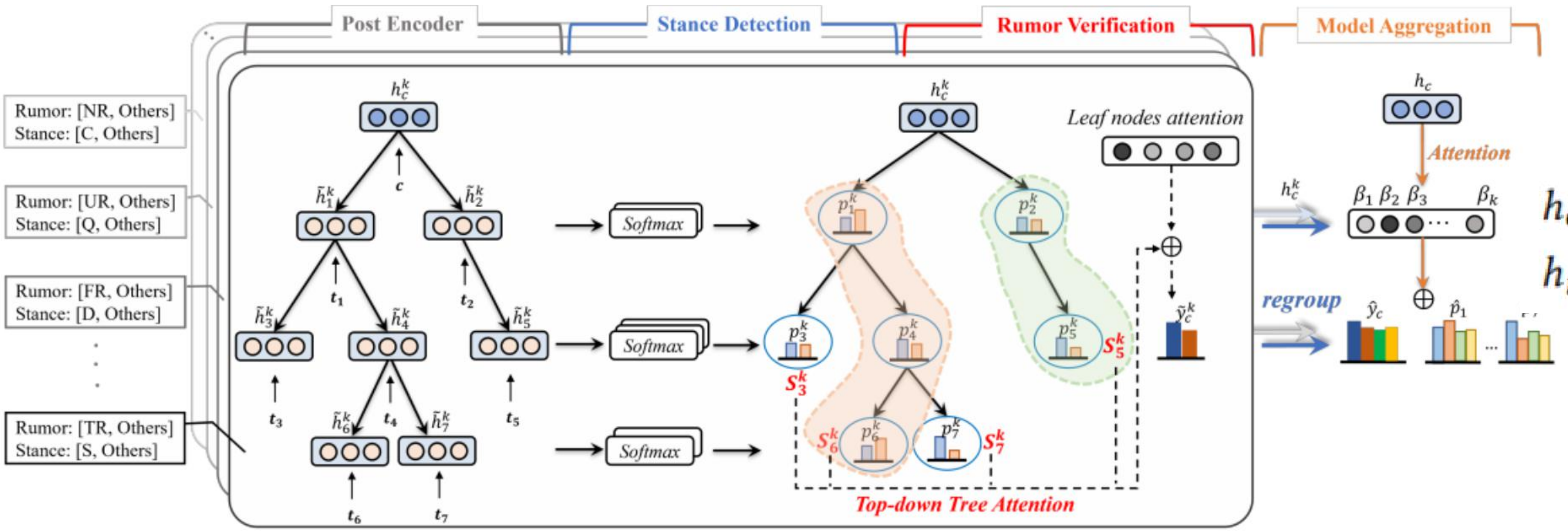


$$C = (c, X, y)$$

Non-rumor (N), True rumor (T), False rumor (F) or Unverified rumor (U)

$$X = (t_1, t_2, \cdots, t_T)$$

$$f(t_1 t_2 \ldots t_T | c) \rightarrow p_1 p_2 \cdots p_T$$

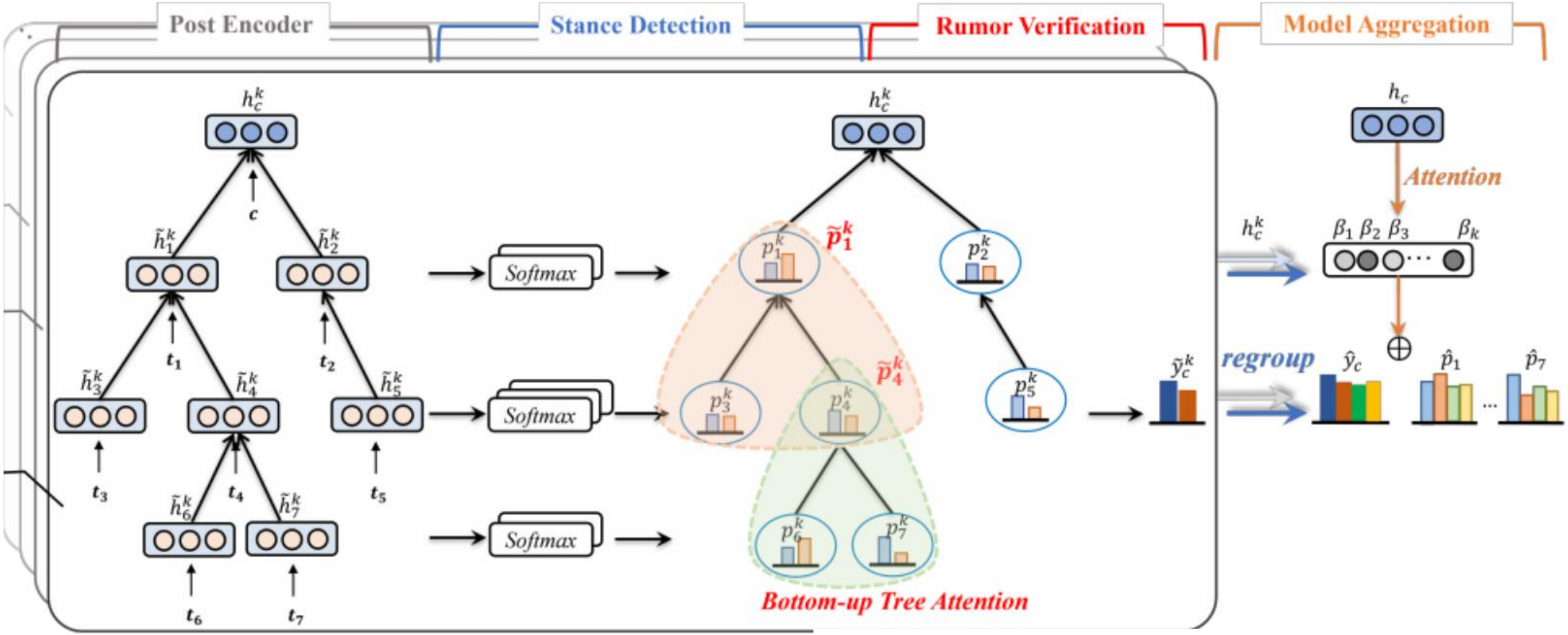


Support (S), Deny (D), Question (Q) or Comment (C)

$$h_c = h_{|c|} = GRU(w_{|c|}, h_{|c|-1}, \theta_c)$$

$$h_i = h_{|t_i|} = GRU(w_{|t_i|}, h_{|t_i|-1}, \theta_X)$$

Method



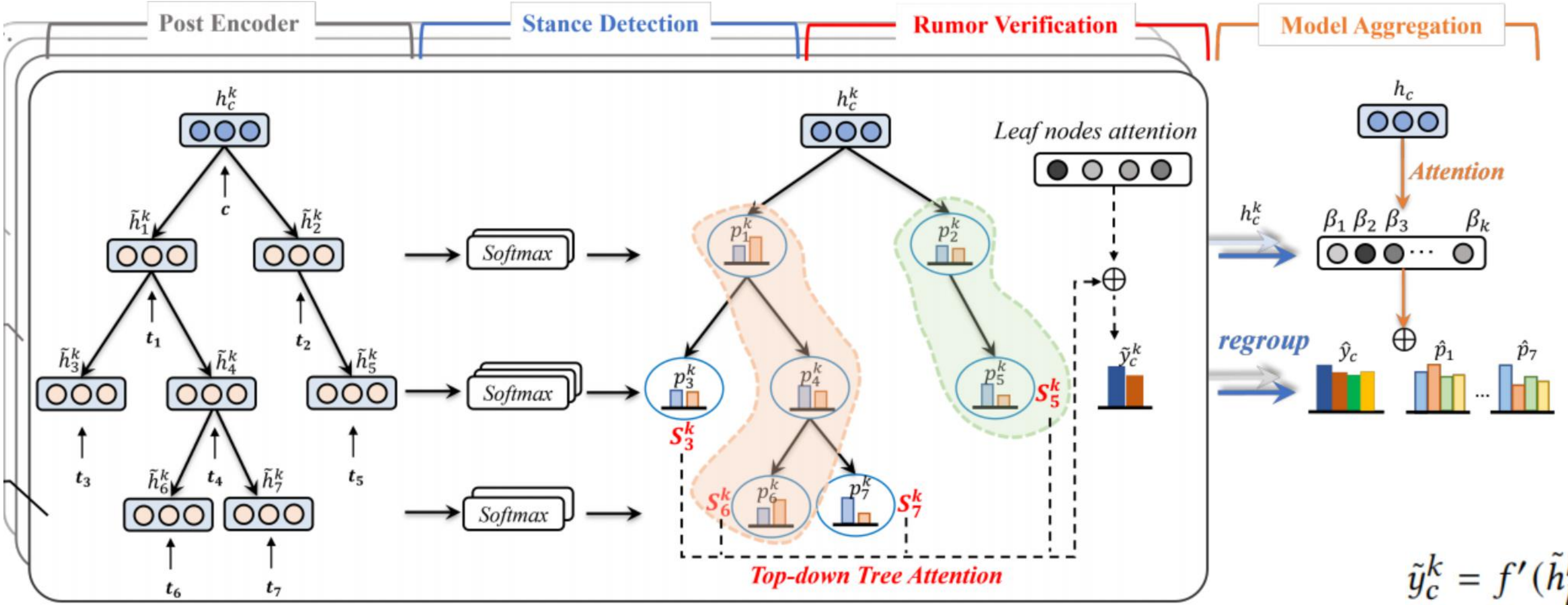
$$\tilde{h}_j^k = RvNN(h_j^k, h_{C(j)}^k, \theta_j^k) \quad (2)$$

$$p_j^k = softmax(W_o^k \tilde{h}_j^k + W_c^k h_c^k + b_o^k) \quad (3)$$

$$\alpha_j^k = \frac{exp(\tilde{h}_j^k \cdot {h_c^k}^\top)}{\sum_{j\in S(i)} exp(\tilde{h}_j^k \cdot {h_c^k}^\top)} \quad (4)$$

$$\tilde{p}_i^k = \sum_{j\in S(i)} \alpha_j^k \cdot \tilde{p}_j^k$$

Method



$$\tilde{y}_c^k = f'(\tilde{h}_l^k, \mathcal{K}(l), s_l^k)$$

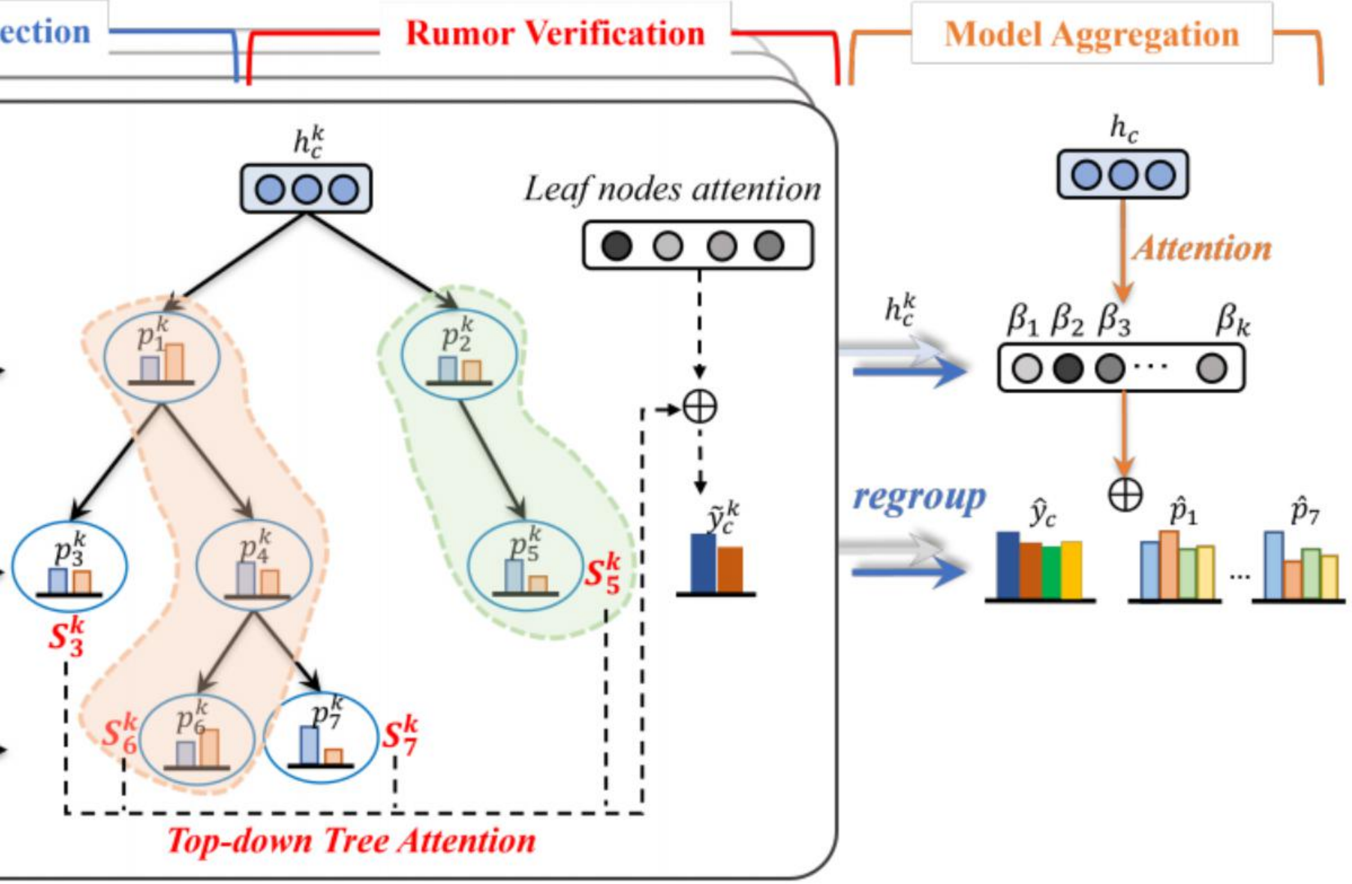
$$\tilde{h}_j^k = RvNN'(h_j^k, h_{P(j)}^k, \theta_j^k) \quad (5)$$

$$p_j^k = softmax(W_o^k \tilde{h}_j^k + W_c^k h_c^k + b_o^k) \quad (6)$$

$$\alpha_j^k = \frac{exp(\tilde{h}_j^k \cdot h_c^{k\top})}{\sum_{i \in \mathcal{P}(l)} exp(\tilde{h}_i^k \cdot h_c^{k\top})} \quad (7)$$

$$s_l^k = \sum_{j \in \mathcal{P}(l)} \alpha_j^k \cdot p_j^k$$

Method



$$h_a = GRU(w_{|c|}, h_{|c|-1}, \theta_a)$$

$$\beta_k = \frac{exp(h_a \cdot {h_c^k}^\top)}{\sum_k exp(h_a \cdot {h_c^k}^\top)} \tag{8}$$

$$\hat{p}_{i,l'} = \sum_{k \in U(l')} \beta_k \cdot p_i^k, \tag{9}$$

where $U(l_s')$ represents the indicator set of the binary classifiers with $l_s' \in [S, D, Q, C]$ as the target class, p_i^k is the predicted stance probability of the post from classifier k, therefore, $\hat{p}_{i,l_s'}$ indicates the probability that the post t_i should be classified as stance l_s'. Thus, the predicted probability distribution over all the stances can be obtained, i.e., $\hat{p}_i = [\hat{p}_{i,S}, \hat{p}_{i,D}, \hat{p}_{i,Q}, \hat{p}_{i,C}]$.

$$\hat{y}_{l_r'} = g'(\beta_k, \tilde{y}^k),$$

$$\hat{y} = [\hat{y}_N, \hat{y}_T, \hat{y}_F, \hat{y}_U]$$

$$y^k = \begin{cases} 1 & \text{if the target of classifier } k \text{ is the same as } y \\ 0 & \text{others} \end{cases} \tag{10}$$

$$L_{bin} = -\sum_{k=1}^{K}\sum_{n=1}^{N} y_n^k * \log \hat{y}_n^k + (1 - y_n^k) * \log(1 - \hat{y}_n^k) \tag{11}$$

$$L_{agg} = -\sum_{n=1}^{N}\sum_{m=1}^{M} y_{m,n} * \log \hat{y}_{m,n} + (1 - y_{m,n}) * \log(1 - \hat{y}_{m,n}) \tag{12}$$

Experiments

Table 1: Statistics of rumor datasets for model training.

Statistics	Twitter15	Twitter16	PHEME
# of claim	1,308	818	6,425
# of Non-rumor	374 (28.6%)	205 (25.1%)	4,023 (62.6%)
# of False-rumor	370 (28.3%)	207 (25.3%)	638 (9.9%)
# of True-rumor	190 (14.5%)	205 (25.1%)	1,067 (16.6%)
# of Unverified-rumor	374 (28.6%)	201 (24.5%)	697 (10.8%)
# tree nodes	68026	40867	383569
# of Avg. posts/tree	52	50	6
# of Max. posts/tree	814	757	228
# of Min. posts/tree	1	1	3

Table 2: Statistics of the datasets for model testing.

Statistics	RumorEval2019-S	SemEval8
# of claim	425	297
# of Non-rumor	100 (23.53%)	——
# of False-rumor	74 (17.41%)	62 (20.8%)
# of True-rumor	145 (34.12%)	137 (46.1%)
# of Unverified-rumor	106 (24.94%)	98 (33.0%)
# posts of Support	1320 (19.65%)	645 (15.1%)
# posts of Deny	522 (7.77%)	334 (7.8%)
# posts of Question	531 (7.90%)	361 (8.5%)
# posts of Comment	4,345 (64.68%)	2,923 (68.6%)
# tree nodes	6,718	4,263
# Avg. posts/tree	16	14
# Max. posts/tree	249	228
# Min. posts/tree	2	3

Experiments

Table 3: Results on stance detection: our methods achieve p-value < 0.05 under t-test for Robustness consideration.

Dataset	RumourEval2019-S							SemEval8						
Method	AUC	MicF	MacF	S F_1	D F_1	Q F_1	C F_1	AUC	MicF	MacF	S F_1	D F_1	Q F_1	C F_1
Zero-Shot	–	0.369	0.324	0.301	0.168	0.342	0.486	–	0.383	0.344	0.278	0.162	0.480	0.456
Pre-Rule	–	0.605	0.478	0.657	0.419	——	——	——	0.429	0.389	0.432	0.644	——	——
C-GCN	0.633	0.629	0.416	0.331	0.173	0.429	0.730	0.610	0.625	0.411	0.327	0.161	0.430	0.728
BrLSTM(V)	0.71	0.66	0.42	0.460	0	0.391	0.758	0.676	0.665	0.401	0.493	0	0.381	0.730
BiGRU(V)	0.7	0.63	0.417	0.392	0.162	0.360	0.754	0.660	0.633	0.416	0.460	0.168	0.328	0.708
MT-GRU(V)	0.714	0.636	0.432	0.313	0.156	0.506	0.748	0.669	0.630	0.413	**0.498**	0.116	0.312	0.729
TD-MIL(V)	0.712	0.65	0.432	0.438	0.156	0.408	0.688	0.668	0.626	0.416	0.473	0.127	**0.463**	0.602
BU-MIL(V)	0.71	0.63	0.431	**0.485**	0.166	0.396	0.688	0.669	0.623	0.415	0.470	0.128	0.460	0.602
TD-MIL(T15)	0.706	0.668	0.427	0.339	0.173	0.444	0.752	0.663	0.642	0.418	0.330	0.174	0.420	0.750
TD-MIL(T16)	0.713	0.665	**0.436**	0.350	**0.182**	0.446	0.758	0.660	**0.671**	0.421	0.334	0.173	0.422	0.754
TD-MIL(PHE)	**0.722**	**0.691**	0.434	0.344	0.179	**0.467**	**0.767**	**0.669**	0.651	**0.426**	0.335	**0.175**	0.430	**0.763**
BU-MIL(T15)	0.706	0.662	0.428	0.341	0.173	0.436	0.756	0.661	0.638	0.415	0.326	0.168	0.420	0.748
BU-MIL(T16)	0.701	0.66	0.426	0.340	0.170	0.438	0.749	0.659	0.637	0.416	0.324	0.169	0.419	0.753
BU-MIL(PHE)	0.707	0.665	0.432	0.344	0.174	0.445	0.762	0.666	0.642	0.420	0.329	0.169	0.423	0.758

Experiments

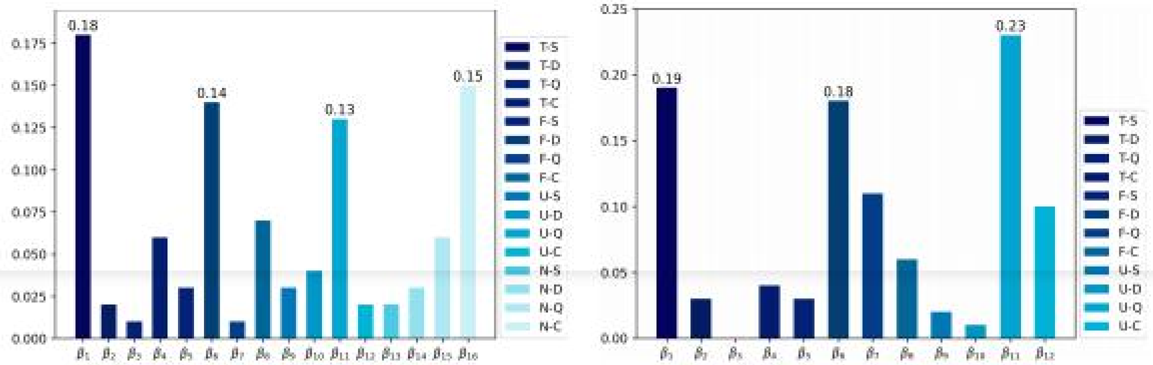
Table 4: Results on Rumor Verification: our methods achieve p-value < 0.05 under t-test for Robustness consideration.

Dataset	RumorEval2019-S							SemEval8					
Method	AUC	MicF	MacF	T F_1	F F_1	U F_1	N F_1	AUC	MicF	MacF	T F_1	F F_1	U F_1
GCAN	0.693	0.645	0.253	0.249	0.31	0.113	0.339	0.688	0.645	0.255	0.241	0.326	0.198
PPC	0.672	0.632	0.25	0.244	0.296	0.114	0.346	0.673	0.642	0.249	0.237	0.289	0.221
TD-RvNN	0.88	0.743	0.699	0.713	0.631	0.660	0.792	0.882	0.728	0.689	0.702	0.619	0.745
BU-RvNN	0.865	0.720	0.723	0.746	0.641	0.696	0.806	0.870	0.708	0.684	0.708	0.620	0.723
H-GCN	0.69	0.534	0.418	0.712	0.180	0.371	0.409	0.675	0.530	0.413	0.355	0.16	0.724
MTL2 (V)	0.683	0.653	0.43	0.622	0.279	0.352	0.457	0.680	0.651	0.433	0.640	0.289	0.372
MT-GRU (V)	0.704	0.768	0.452	0.462	0.298	0.373	0.452	0.701	0.761	0.428	0.639	0.254	0.391
TD-MIL (V)	0.685	0.678	0.45	0.667	0.329	0.376	0.428	0.680	0.621	0.436	0.650	0.274	0.384
BU-MIL (V)	0.682	0.679	0.448	0.668	0.326	0.373	0.428	0.680	0.645	0.427	0.631	0.292	0.360
TD-MIL (T15)	**0.919**	0.793	**0.79**	0.822	**0.762**	0.716	0.818	**0.913**	0.771	0.730	0.679	**0.689**	0.823
TD-MIL (T16)	0.914	0.792	0.764	0.796	0.740	0.719	0.812	0.899	0.785	0.725	0.668	0.682	0.825
TD-MIL (Phe)	0.917	**0.809**	0.776	**0.826**	0.659	0.669	**0.852**	0.908	**0.798**	**0.741**	**0.741**	0.672	0.810
BU-MIL (T15)	0.899	0.769	0.78	0.794	0.688	0.770	0.819	0.887	0.752	0.724	0.670	0.680	0.822
BU-MIL (T16)	0.902	0.776	0.76	0.780	0.664	**0.780**	0.810	0.893	0.756	0.721	0.663	0.676	**0.826**
BU-MIL (Phe)	0.904	0.776	0.763	0.793	0.666	0.770	0.833	0.902	0.763	0.729	0.728	0.649	0.809

Experiments

Table 5: Ablation Study Results

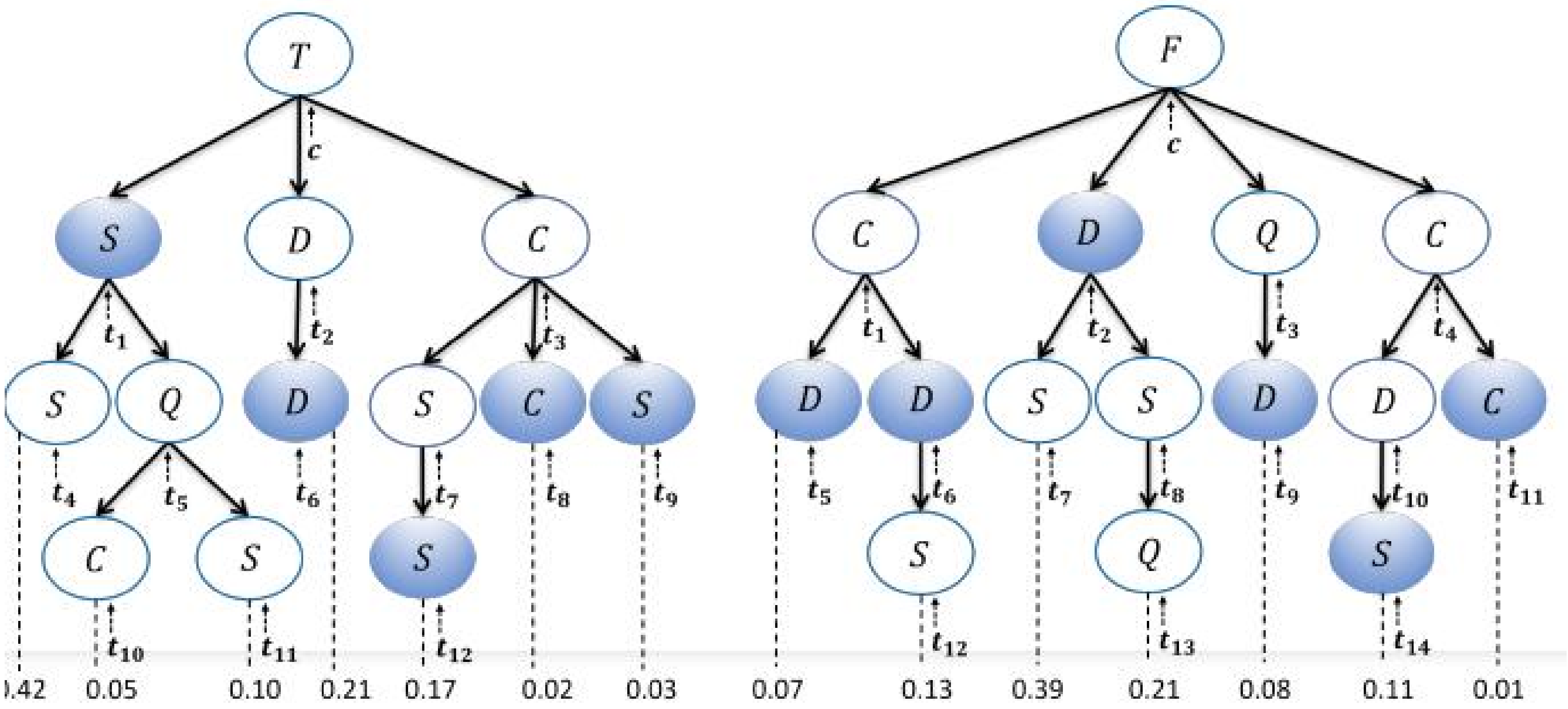
	Rumor Result			Stance Result		
Method	AUC	MicF	MacF	AUC	MicF	MacF
MIL-a	0.892	0.759	0.736	0.672	0.643	0.43
TD-MIL-b	0.912	0.802	0.746	0.701	0.658	0.426
TD-MIL-c	0.903	0.805	0.738	0.696	0.653	0.42
BU-MIL-b	0.901	0.752	0.743	0.698	0.647	0.419
BU-MIL-c	0.903	0.749	0.742	0.687	0.645	0.419
TD-MIL	0.917	0.809	0.776	0.722	0.691	0.434
BU-MIL	0.904	0.776	0.763	0.707	0.665	0.432



(a) RumorEval2019-S (b) SemEval8

Figure 5: Average Attention Score for Binary Classifiers from Eq. 8). RumorEval2019-S dataset has 16 binary classifiers and SemEval8 dataset has 12 binary classifiers.

Experiments



(a) True rumor case (b) False rumor case

Figure 4: Case Study for Tree Attention Mechanism.

Thanks