Hashtag-Guided Low-Resource Tweet Classification

Shizhe Diao*
The Hong Kong University of Science and Technology
sdiaoaa@connect.ust.hk

Sedrick Scott Keh*
Carnegie Mellon University
skeh@cs.cmu.edu

Liangming Pan
University of California, Santa Barbara
liangmingpan@ucsb.edu

Zhiliang Tian
The Hong Kong University of Science and Technology
tianzhilianghit@gmail.com

Yan Song
University of Science and Technology of China
clksong@gmail.com

Tong Zhang
The Hong Kong University of Science and Technology
tongzhang@ust.hk

WWW2023

code: https://github.com/shizhediao/HashTation.

Reported by Xiaoke Li

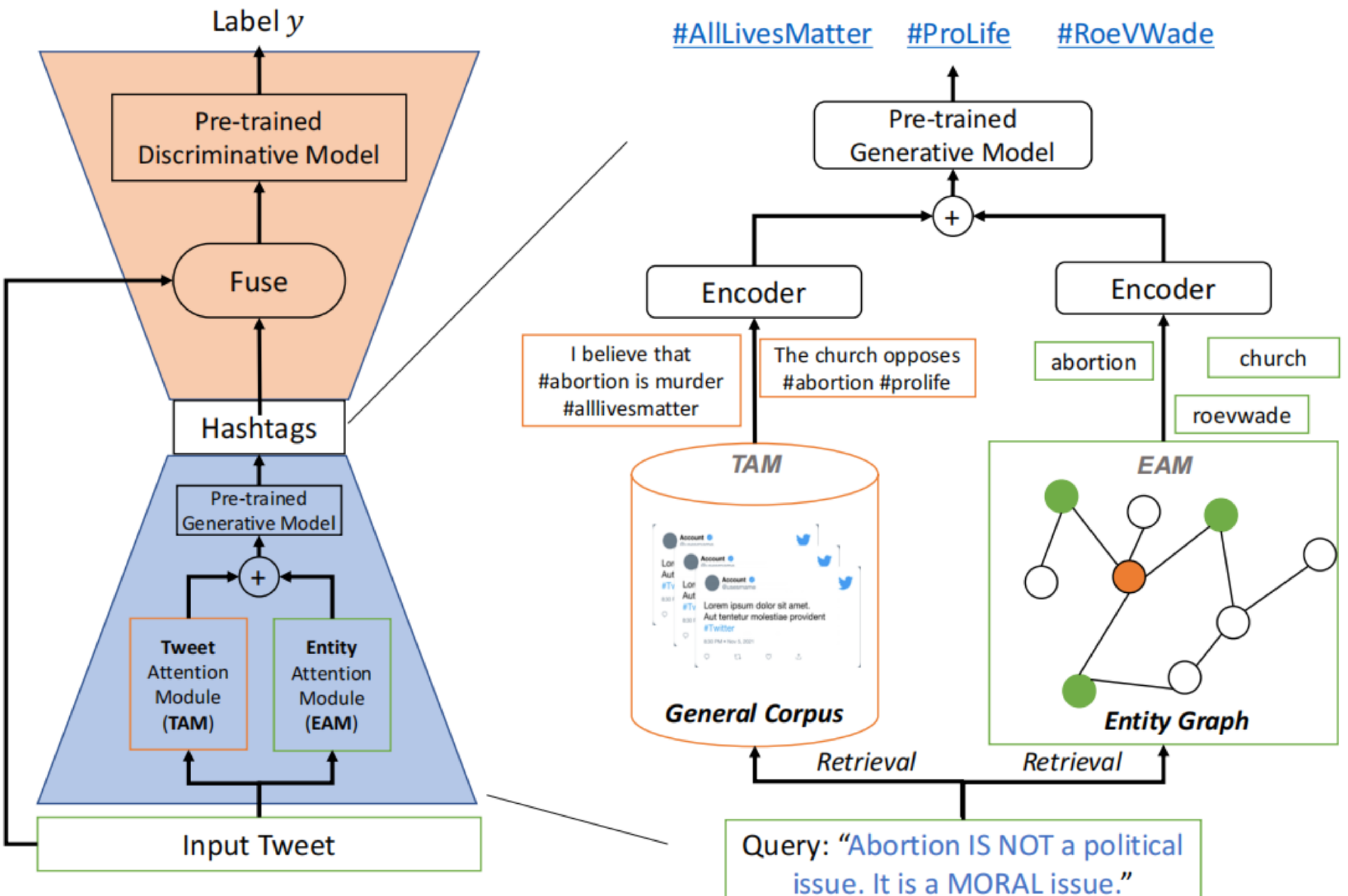
Introduction

Table 1: Examples of how hashtags can provide auxiliary information for better tweet classification.

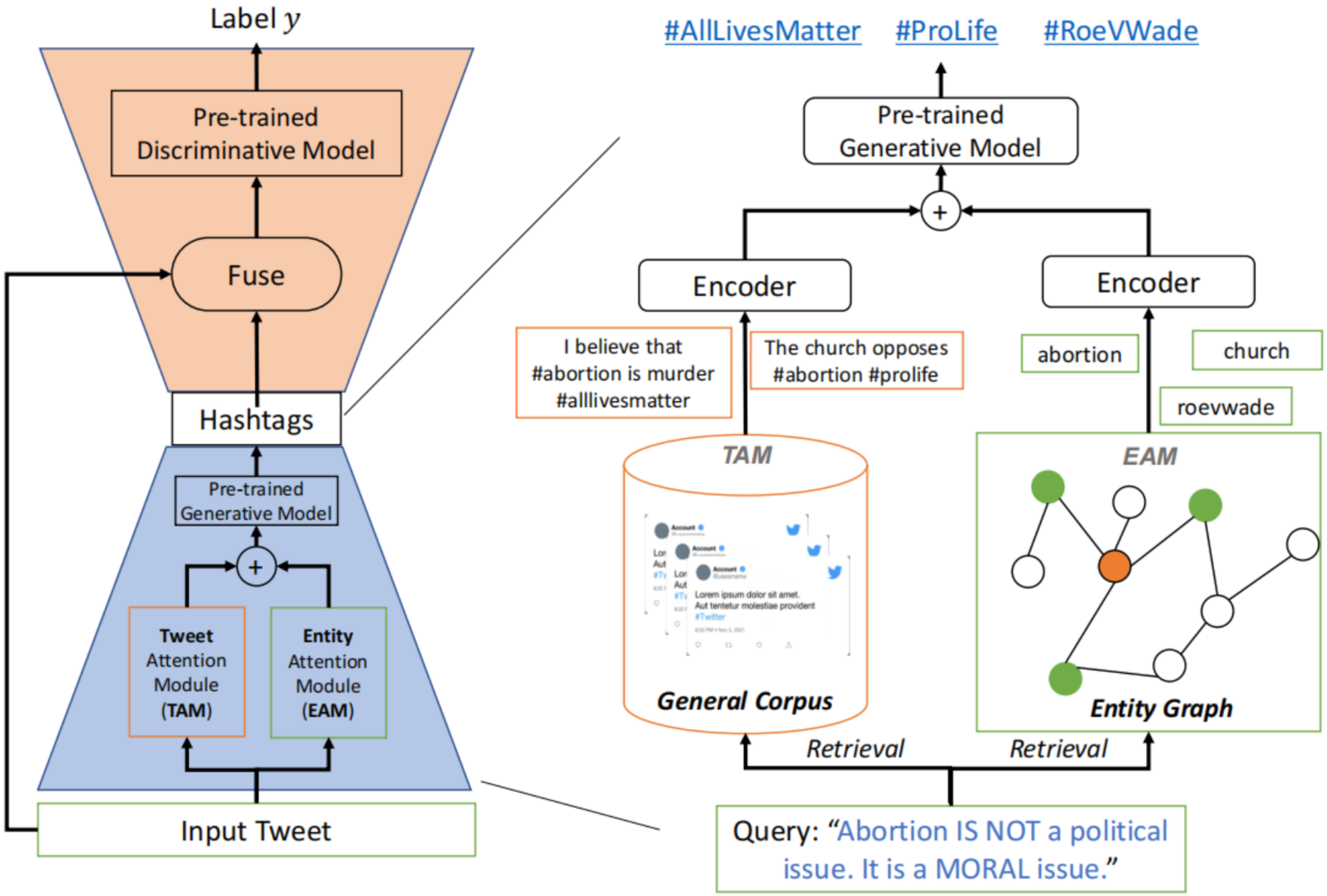
Original Input	Generated Hashtags	New Hashtag-Guided Input
Abortion IS NOT a political issue. It is a MORAL issue	AllLivesMatter , ProLife	Abortion IS NOT a political issue. It is a MORAL issue #AllLivesMatter #ProLife
Twitter making everybody mad. It's hilarious	Twitter, hilarious	#Twitter making everybody mad. It's #hilarious
He's the GOAT for sure!	GOAT, NBAFinals	He's the #GOAT for sure! #NBAFinals

	EMOJI	EMOTION	HATE	IRONY	OFFENSIVE	SENTIMENT	STANCE	AVG.
w.o. HASHTAGS	8.26	55.9	63.7	61.7	64.1	50.2	55.5	51.3
w. HASHTAGS	8.58	64.5	73.5	67.3	65.8	52.0	59.4	55.9

Method


Label y
Pre-trained Discriminative Model
Fuse
Hashtags
Pre-trained Generative Model
+
Tweet Attention Module (TAM)
Entity Attention Module (EAM)
Input Tweet
#AllLivesMatter
#ProLife
#RoeVWade
Pre-trained Generative Model
+
Encoder
Encoder
I believe that #abortion is murder #alllivesmatter
The church opposes #abortion #prolife
abortion
church
roevwade
TAM
General Corpus
EAM
Entity Graph
Retrieval
Retrieval
Query: "Abortion IS NOT a political issue. It is a MORAL issue."

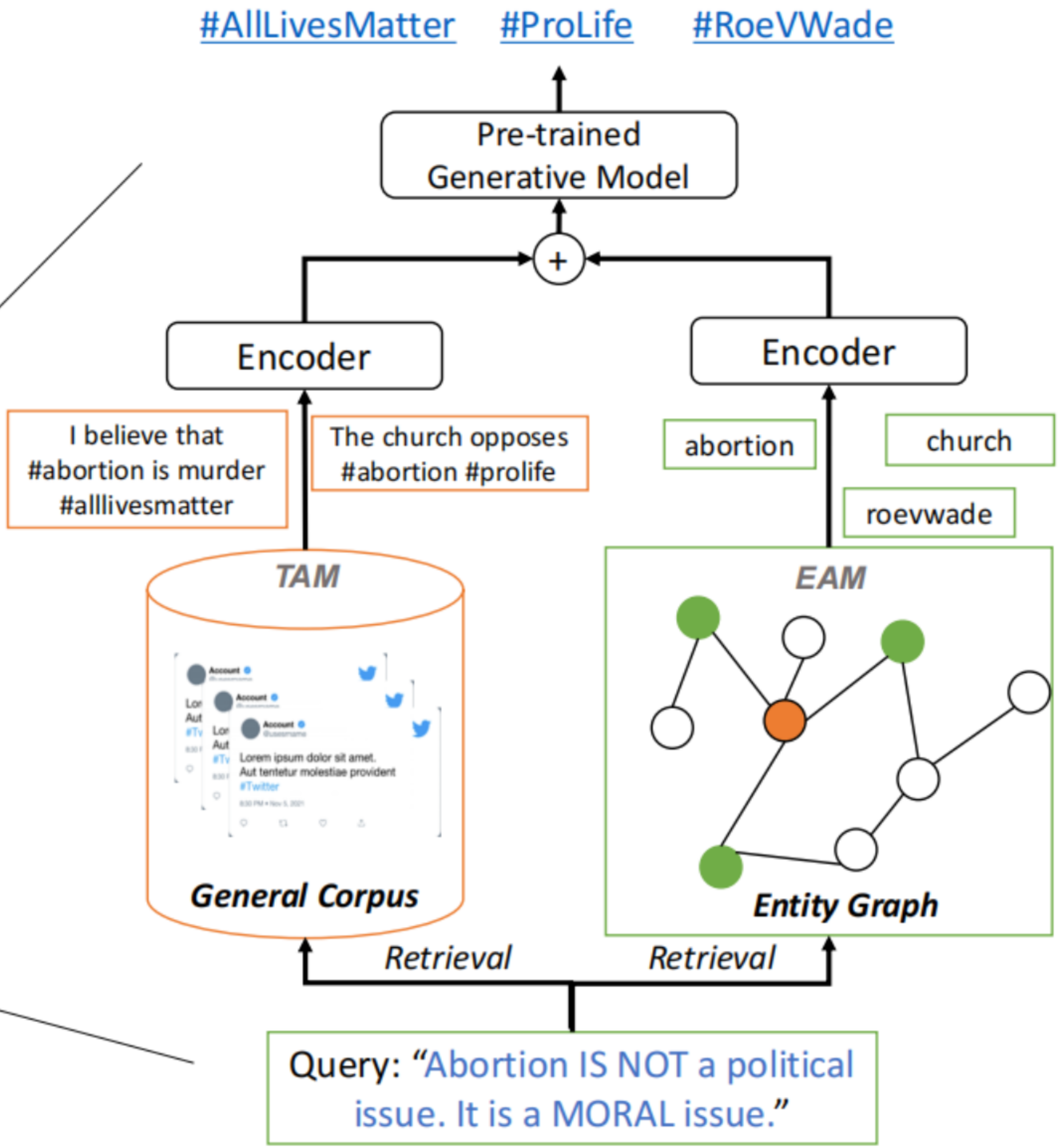
Method



$$H = \text{HASH-GEN}(x, \text{TAM}(x, \mathcal{D}), \text{EAM}(x, \mathcal{D})), \quad (1)$$

$$y = \text{TWEET-CLASSIFIER}(\mathit{Fuse}(H, x)), \quad (2)$$

Method



$$\mathcal{D} = \{x_1, ..., x_k, ..., x_l\}$$

$$\text{key vectors } \mathcal{U} = \{\boldsymbol{u}_1, ..., \boldsymbol{u}_k, ..., \boldsymbol{u}_l\}$$

$$\text{value vectors } \mathcal{V} = \{\boldsymbol{v}_1, ..., \boldsymbol{v}_k, ..., \boldsymbol{v}_l\}$$

$$p_k = \frac{\exp(\boldsymbol{e}^\top \cdot \boldsymbol{u}_k)}{\sum_{k=1}^{l} \exp(\boldsymbol{e}^\top \cdot \boldsymbol{u}_k)}, \tag{3}$$

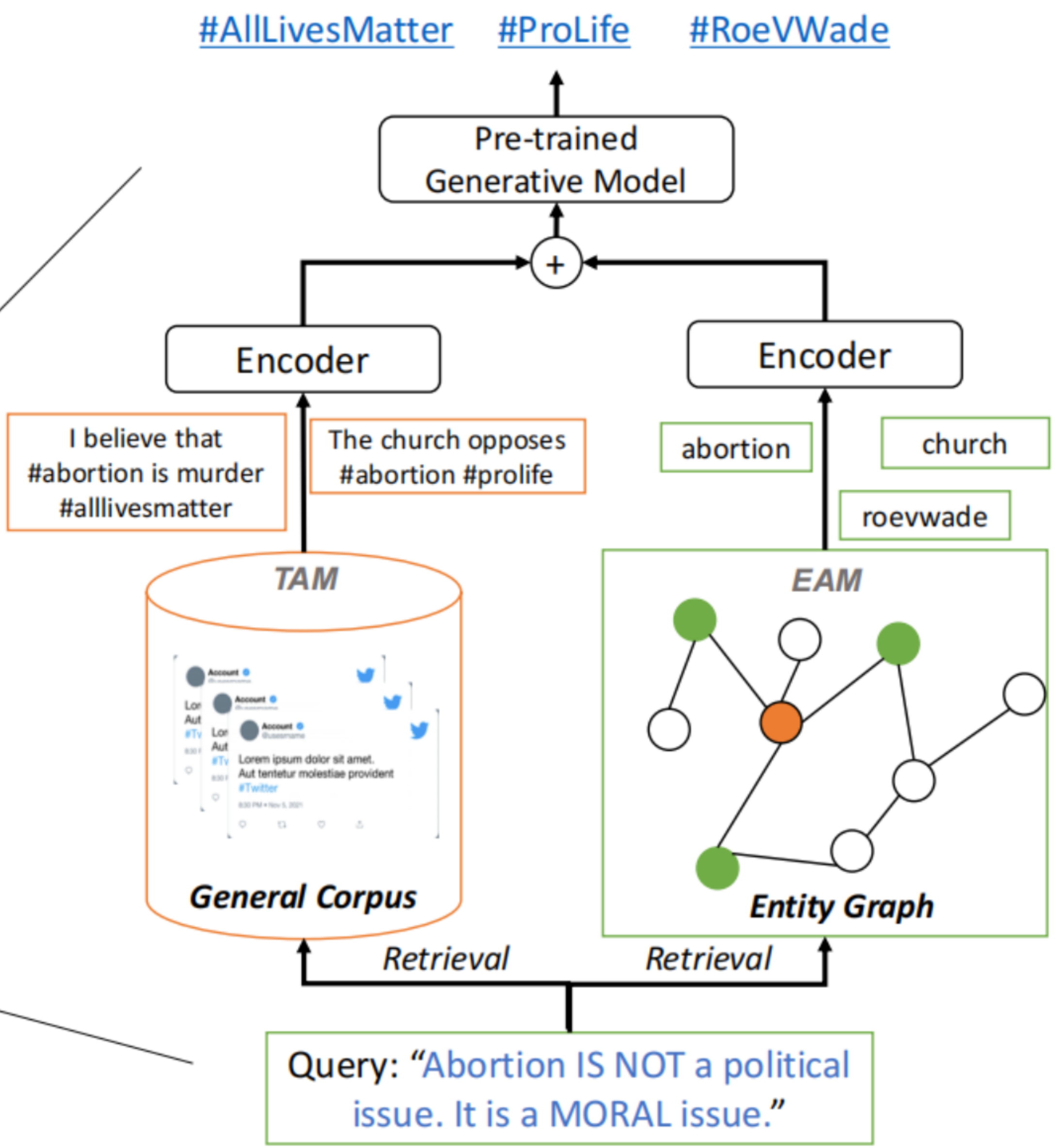
and for the entire set $\mathcal{D}$, we have $\boldsymbol{o} = \sum_{k=1}^{l} p_k \boldsymbol{v}_k$, where $\boldsymbol{o}$ is the output vector of the TAM to represent the latent topics from relevant tweets via a weighted encoding.

$$H^{(l+1)} = \sigma\left(\hat{D}^{(-\frac{1}{2})} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}\right), \tag{4}$$

$$\alpha_{ij} = \frac{\exp\left([\mathbf{W}^Q \vec{\boldsymbol{h}}_i^l \mathbf{W}^K \vec{\boldsymbol{h}}_j^l]\right)}{\sum_{k \in M_i} \exp\left([\mathbf{W}^Q \vec{\boldsymbol{h}}_i^l \mathbf{W}^K \vec{\boldsymbol{h}}_k^l]\right)}, \tag{5}$$

$$\vec{\boldsymbol{h}}_i^{l+1} = \sigma\left(\sum_{n_j \in M_i} \alpha_{ij} \mathbf{W} \vec{\boldsymbol{h}}_j^l\right), \tag{6}$$

Method



$$\tilde{\boldsymbol{h}}_i = \boldsymbol{h}_i + \boldsymbol{o}.$$

$$\tilde{\boldsymbol{h}}_i = \tilde{\boldsymbol{h}}_i + \sum_j \vec{\boldsymbol{h}}_{i,j}$$

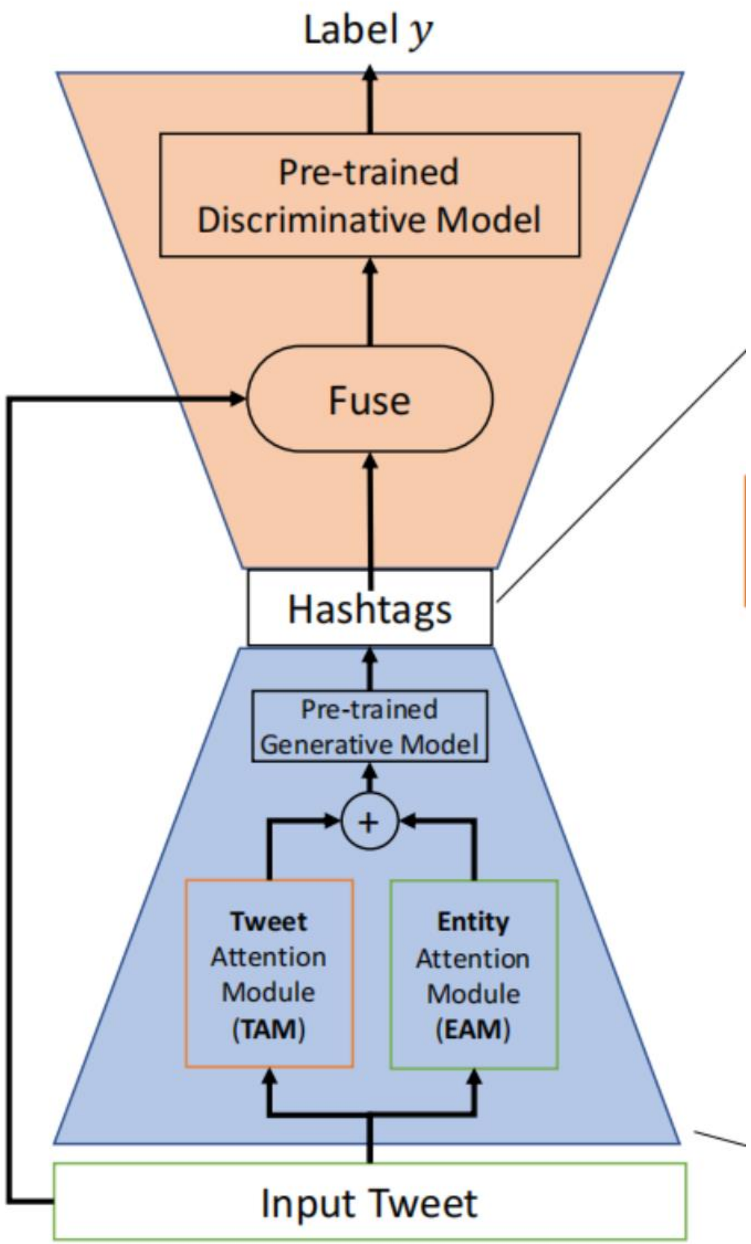
$$\boldsymbol{a}^t = \alpha_1^t \ldots \alpha_i^t \ldots \alpha_n^t$$

$$\boldsymbol{c}^t = \sum_{i=1}^n \alpha_i^t \tilde{\boldsymbol{h}}_i$$

$$\boldsymbol{d}_v = \frac{\exp(\boldsymbol{z}^t)}{\sum_V \exp(\boldsymbol{z}^t)}, \tag{7}$$

where $\boldsymbol{z}^t = \boldsymbol{W}_1(\boldsymbol{W}_2 \cdot (\boldsymbol{s}^t \oplus \boldsymbol{c}^t))$, a vector with $|V|$ dimension and V is the predefined vocabulary providing word candidates for hashtag generation. $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are trainable parameters for the two aforementioned linear layers, respectively.

Method



After obtaining the desired hashtags, we fuse the hashtags $H = h_1, h_2, .., h_n$ with the input tweet x to obtain the hashtag-guided tweet $Fuse(H, x)$. There are several ways to implement the $Fuse$ function like simple concatenation, prompting with a manual template, and so on. We explore the effects of different strategies in Section 5.1. Finally, we train a tweet classifier $\mathcal{F}$ with the following objective:

$$\min_{\phi}(\mathcal{L}(\mathcal{F}(Fuse(H, x)), Y)), \tag{8}$$

where $\mathcal{L}$ is the cross-entropy loss function following the standard practice in tweet classification [30, 38].

Experiments

Table 3: Dataset statistics of the original dataset and the set whose tweets contain hashtags.

Task	Labels	Original Dataset |Train|	Original Dataset |Val|	Original Dataset |Test|	With Hashtags |Train|	With Hashtags |Val|	With Hashtags |Test|
Emoji Prediction	20 different emojis	44489	4930	49664	19878	1800	20387
Emotion Recognition	anger, joy, sadness, optimism	3188	363	1384	1394	178	650
Hate Speech Detection	hateful, not hateful	8914	990	2963	2223	228	1394
Irony Detection	irony, not irony	2802	942	781	1017	368	530
Offensive Language Identification	offensive, not offensive	11616	1301	852	1664	194	630
Sentiment Analysis	positive, neutral, negative	45612	2000	12200	8262	347	4571
Stance Detection	in favor, neutral, against	2620	294	1249	2578	288	1249

Experiments

Table 4: F1-scores for baseline models (top) and our proposed method and its variants (bottom). Best results are highlighted in bold. We report the average score over ten different random seeds. * denotes HASHTATION has significant differences ($p-value$ <0.05) over baseline models.

	EMOJI	EMOTION	HATE	IRONY	OFFENSIVE	SENTIMENT	STANCE	AVG.
KIM-CNN	3.8	20.1	51.1	39.8	41.9	39.5	32.2	32.6
BILSTM	5.9	26.8	49.2	37.4	47.6	43.1	28.7	34.1
BERT-BASE	11.2	57.7	51.6	53.7	56.4	56.6	52.3	48.5
ROBERTA-BASE	11.8	58.9	56.7	54.2	59.7	57.3	54.1	50.4
BERTWEET	12.1	59.4	55.5	57.0	61.8	59.0	55.5	51.5
HASHTATION-BT	12.7	61.0	56.4	58.6	**63.8***	**59.9***	56.3	52.7
W/O TAM	12.5	60.6	56.3	57.7	63.0	59.8	56.2	52.3
W/O EAM	12.6	60.5	55.7	58.6	63.5	59.4	56.3	52.4
TIMELMS	12.4	60.2	56.9	59.6	60.0	57.4	55.9	51.8
HASHTATION-TL	**13.0***	**61.6***	**58.0***	**60.9***	62.3	59.6	**56.5***	**53.1***
W/O TAM	12.7	61.2	57.6	60.6	61.9	58.9	56.1	52.7
W/O EAM	12.7	61.3	57.6	60.7	61.6	59.0	56.2	52.7

Experiments

Table 5: Effects of different fusion methods on the TimeLMS model. We explore four methods: no hashtags, standard, pre-pending at the start, appending at the end. F1-scores are reported on seven tasks.

Fusion Method	Emoji	Emotion	Hate	Irony	Offensive	Sentiment	Stance
Without Hashtags	12.4	60.2	56.9	59.6	60.0	57.4	55.9
Standard	**13.0**	61.6	**58.0**	**60.9**	**62.3**	**59.6**	56.5
Start	12.9	61.0	**58.0**	60.7	61.3	59.3	56.1
End	**13.0**	**61.7**	57.6	60.5	61.9	59.6	**56.6**

Experiments

Table 6: Examples of hashtags generated by HASHTATION, as compared to ground truth hashtags.

Text	Dataset	Classification	Ground Truth Hashtags	Generated Hashtags
Riding with @user and some incredible people. Truly magical.	Emoji	sparkles	#soulcycle #lululemon	#soulcycle
Oh Canada shouldn't be sung like that.	Emotion	Anger	#terrible #MLBTHESHOW17	#terrible #Canada
Both #Frightening and #Demented #Sick. Now it will be our problem.	Hate	No	#Frightening, #Demented, #Sick, #SendThemBack, #KAG	#Sick #BuildThatWall #MAGA #KAG
Oh how I love working in Baltimore	Irony	Yes	#not	#not #Baltimore
@user Ya Obama on trade 2-YEARS AGO: "Trump is just NOT TELLING THE TRUTH How STUPID could our leaders be" God bless Trump!	Offensive	Yes	#MAGA #KAG	#MAGA #Trump
Messi: "To have a good team, everything starts there"	Sentiment	Joy	#fcblive	#fcblive
"Manspreading"? But women hog subway space, too!	Stance (Feminism)	against	#doublestandards	#doublestandards #manspreading

Experiments

Table 7: Average precision, recall, and F1 scores for the hashtag generation task across all datasets.

Type	Precision	Recall	F1-Score
Bi-Attn [51]	0.202	0.069	0.103
One2Set [56]	0.231	0.095	0.134
SEGTRM [32]	**0.352**	0.127	0.187
HASHTATION (ours)	0.340	**0.135**	**0.193**
Present	0.848	0.234	0.381
Absent	0.171	0.108	0.122

Experiments

Table 8: An example of an input text whose stance is not very obvious and is implied instead of explicitly stated. We highlight the corresponding entities and relevant tweets that are extracted by the EAM and TAM modules.

(a) Example Input Text and Corresponding Hashtags (Entities from EAM highlighted in blue)

Input Text	G. Truth Hashtags	Pred. Hashtags
3 cases of COVID-19 (coronavirus) in 2 schools in my city, both involving teachers coming back from the north of Italy and having contact with the children for almost a week before anything was done.	#coronavirusuk, #CloseTheSchools	#CloseTheSchools

(b) Tweets with Relevant Entities (extracted by EAM)

What a lack of intestinal fortitude MENTION you will have the deaths of Australians on your hands. When we reach 12 pages of obituaries like Italy please look in the mirror for the cause.	#lockusdown, #coronavirus, #CloseTheSchools
At the end of the day, schools with sealed windows, and interior classrooms will have Coronavirus buildup that will increase COVID19 viral load! These building shouldn't be used. Children and teachers are not going to be victims of "risk mitigation"	#Coronavirus, #COVID19, #NotMyChild

(c) Most Relevant Tweets (extracted and ranked by TAM)

We would all feel very different about schools reopening if we had a goverment that: -Was trustworthy -Had best interest of children as only motivation -Based on Science -provided necessary resources to ALL schools -had a National plan None of these is true	#NotMyChild
Schools should just be closed,the 2020 curriculum will continue once the vaccine is found, even if it means 2020 curriculum get done in 2021-2022, we'll have to adjust and catch-up at some point.	#SchoolsMustShutDown

Thanks