A Supplemental Materials

A.1 Parameters of Embedding Pre-training

Word embeddings were pre-trained with the parameters of learning rate=0.025, window size=5, negative sample size=5, and epoch=5. User embeddings were pre-trained with the parameters of initial learning rate=0.025, order=2, negative sample size=5, and training sample size=100M.

A.2 Model Parameters and Parameter Selection Strategies

Unit Sizes, Embedding Dimensions, and a Max Tweet Number

The layers and the embeddings in our models have unit size and embedding dimension parameters. We also restricted the maximum number of tweets per user for TwitterUS to reduce memory footprints. Table 5 shows the values for these parameters. Smaller values were set for TwitterUS because TwitterUS is approximately 2.6 times larger in terms of tweet number. It was computationally expensive to process TwiiterUS in the same settings as W-NUT.

Regularization Parameters and Bucket Sizes

We chose optimal values of α using a grid search with the development sets of TwitterUS and W-NUT. The range of α was set as the following:
$\alpha \in \{1e^{-4}, 5e^{-5}, 1e^{-5}, 5e^{-6}, 1e^{-6}, 5e^{-7}, 1e^{-7}, 5e^{-8}, 1e^{-8}\}$.

We also chose optimal values of c using grid search with the development sets of TwitterUS and W-NUT for the baseline models. The range of c was set as the following for TwitterUS:
$c \in \{50, 100, 150, 200, 250, 300, 339\}$.
The following was set for W-NUT:
$c \in \{100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3028\}$.
Table 6 presents selected values of α and c. For LR-STACK and MADCEl-B-LR-STACK, different parameters of α and c were selected for each logistic regression classifier.

MAD Parameters and Celebrity Threshold

The MAD parameters μ_1, μ_2, and μ_3 and celebrity threshold t were also chosen using grid search with the development sets of TwitterUS and W-NUT. The ranges of μ_1, μ_2, and μ_3 were set as the following:
$\mu_1 \in \{1.0\}$, $\mu_2 \in \{0.001, 0.01, 0.1, 1.0, 10.0\}$,
$\mu_3 \in \{0.0, 0.001, 0.01, 0.1, 1.0, 10.0\}$.
The range of t for TwitterUS was set as $t \in \{2, \ldots, 16\}$. The range of t for W-NUT was set as $t \in \{2, \ldots, 6\}$. Table 6 presents selected values of μ_1, μ_2, μ_3, and t.

	TwitterUS	W-NUT
RNN unit size	100	200
Attention context vector size	200	400
FC unit size	200	400
Word embedding dimension	100	200
Timezone embedding dimension	200	400
City embedding dimension	200	400
User embedding dimension	200	400
Max tweet number per user	200	-

Table 5: Unit sizes, embedding dimensions, and max tweet numbers of our models.

Model	Parameter	TwitterUS	W-NUT
SUB-NN-TEXT	α	$1e^{-8}$	$1e^{-7}$
SUB-NN-UNET	α	$1e^{-6}$	$5e^{-8}$
SUB-NN-META	α	$1e^{-8}$	$5e^{-8}$
Proposed Model	α	$1e^{-6}$	$5e^{-8}$
LR	α	$1e^{-6}$	$5e^{-7}$
MADCEL-B-LR	c	300	3000
	α_{MSG}	$1e^{-6}$	$5e^{-7}$
	α_{LOC}	$1e^{-6}$	$1e^{-6}$
	α_{DESC}	$5e^{-6}$	$1e^{-6}$
	α_{TZ}	$1e^{-4}$	$5e^{-6}$
LR-STACK	α_{2ND}	$1e^{-6}$	$1e^{-7}$
MADCEL-B-LR-STACK	c_{MSG}	300	3000
	c_{LOC}	300	3000
	c_{DESC}	250	1500
	c_{TZ}	100	2500
	c_{2ND}	300	2000

Table 6: Regularization parameters and bucket sizes selected for our models and baseline models.

Model	Parameter	TwitterUS	W-NUT
MADCEL-B-LR	μ_1	1.0	1.0
	μ_2	1.0	10.0
	μ_3	0.01	0.1
	t	5	4
MADCEL-B-LR-STACK	μ_1	1.0	1.0
	μ_2	1.0	1.0
	μ_3	0.1	0.0
	t	4	2

Table 7: MAD parameters and celebrity threshold selected for baseline models.