Unifying Text, Metadata, and User Network Representations with a Neural Network for Geolocation Prediction

Yasuhide Miura[†,‡], Motoki Taniguchi[†],

Tomoki Taniguchi[†], and Tomoko Ohkuma[†]

[†] Fuji Xerox Co., Ltd.

[‡] Tokyo Institute of Technology

Geolocation Prediction

- Goal
  – Predict a location of a person

Example

My house is at Vancouver.

an SNS message

geolocation prediction

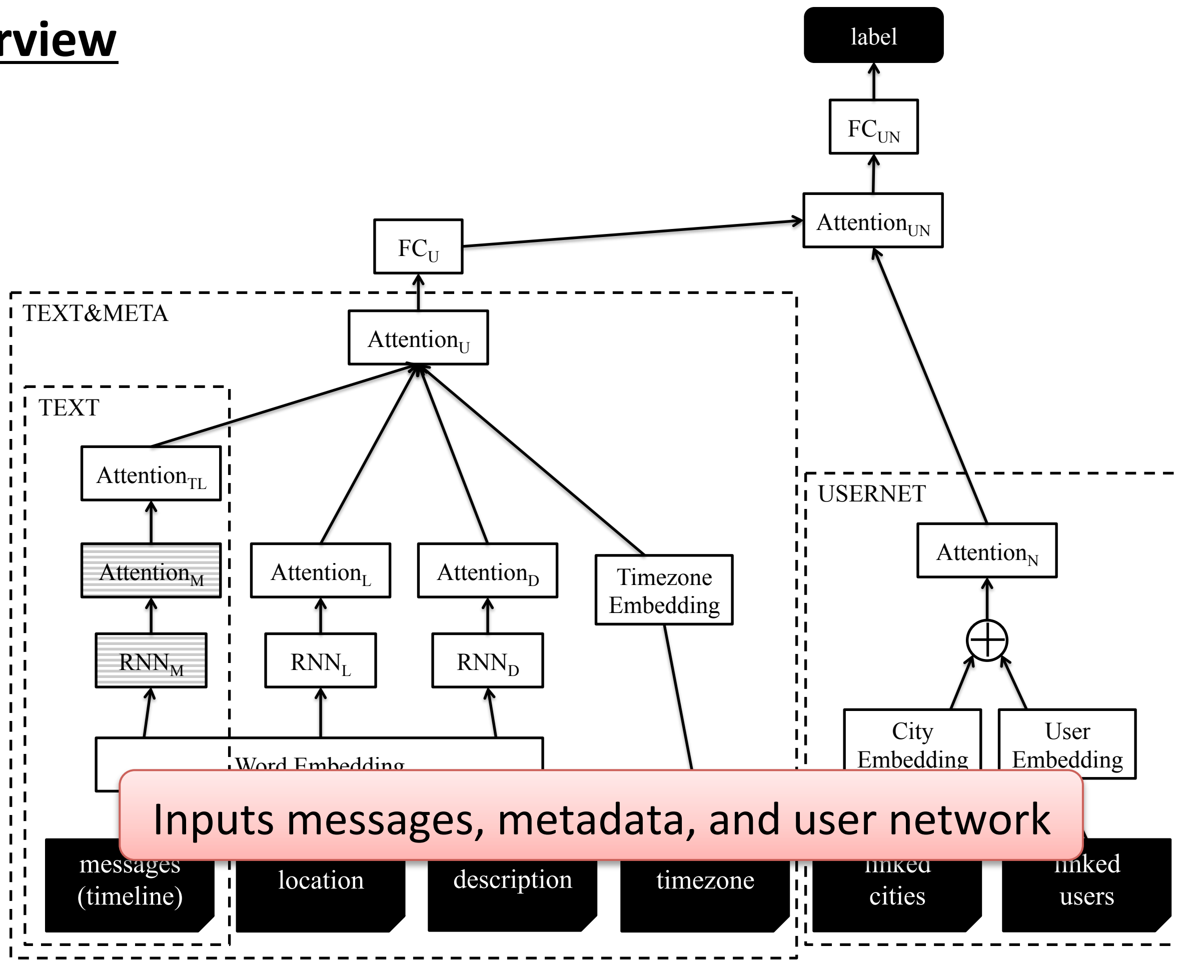
Vancouver

city name

Our Approach

- Geolocation prediction with neural networks

Overview

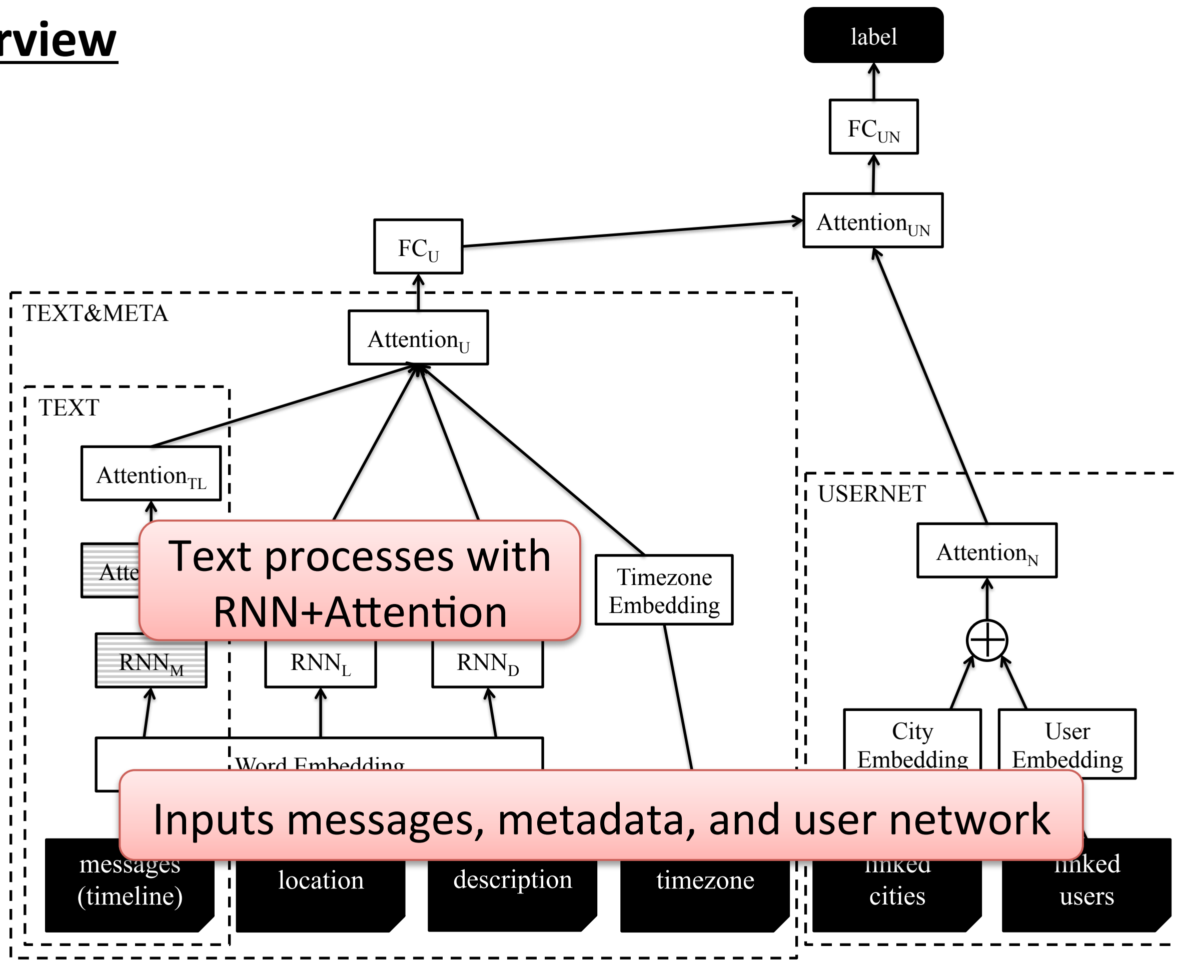


Inputs messages, metadata, and user network

Our Approach

- Geolocation prediction with neural networks

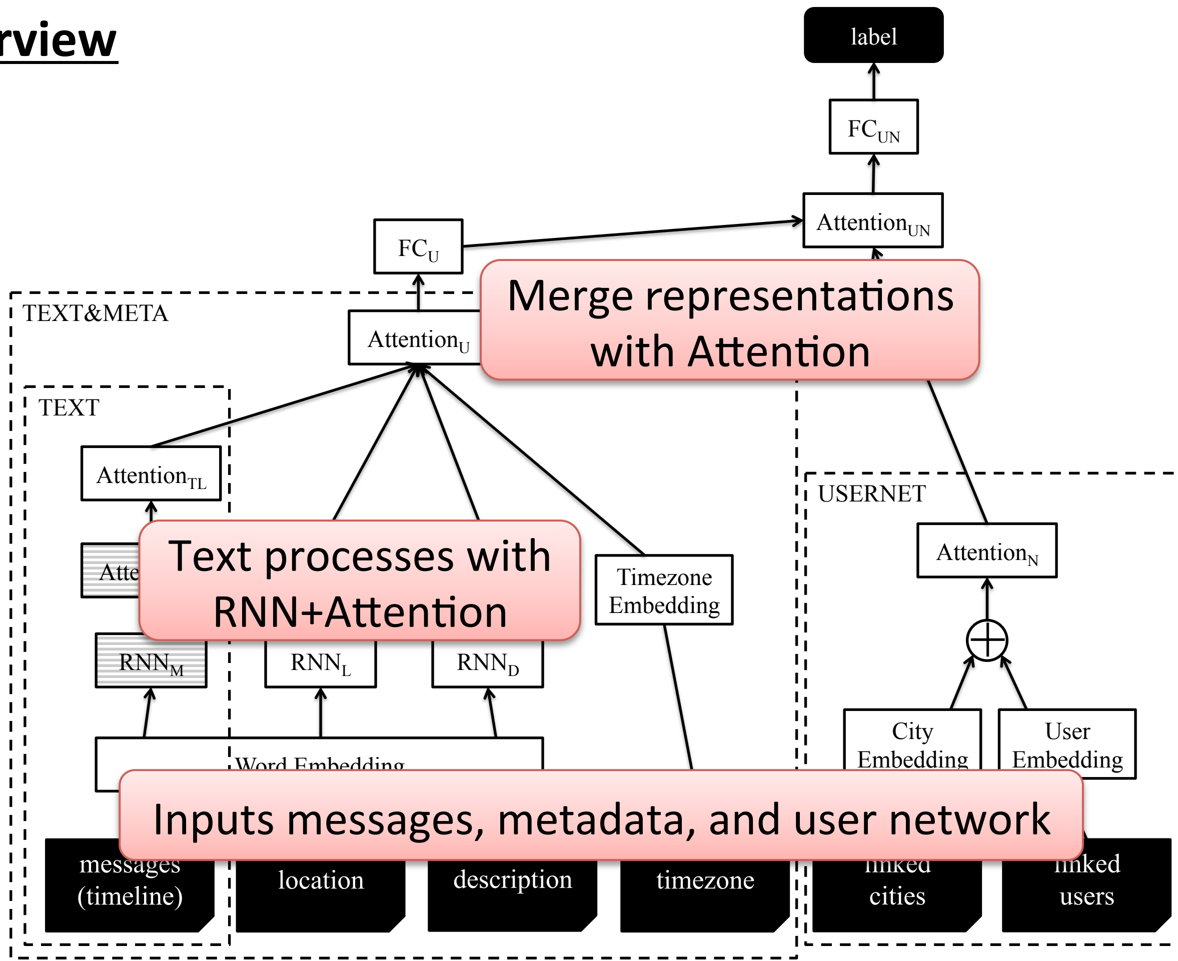
Overview



Our Approach

- Geolocation prediction with neural networks

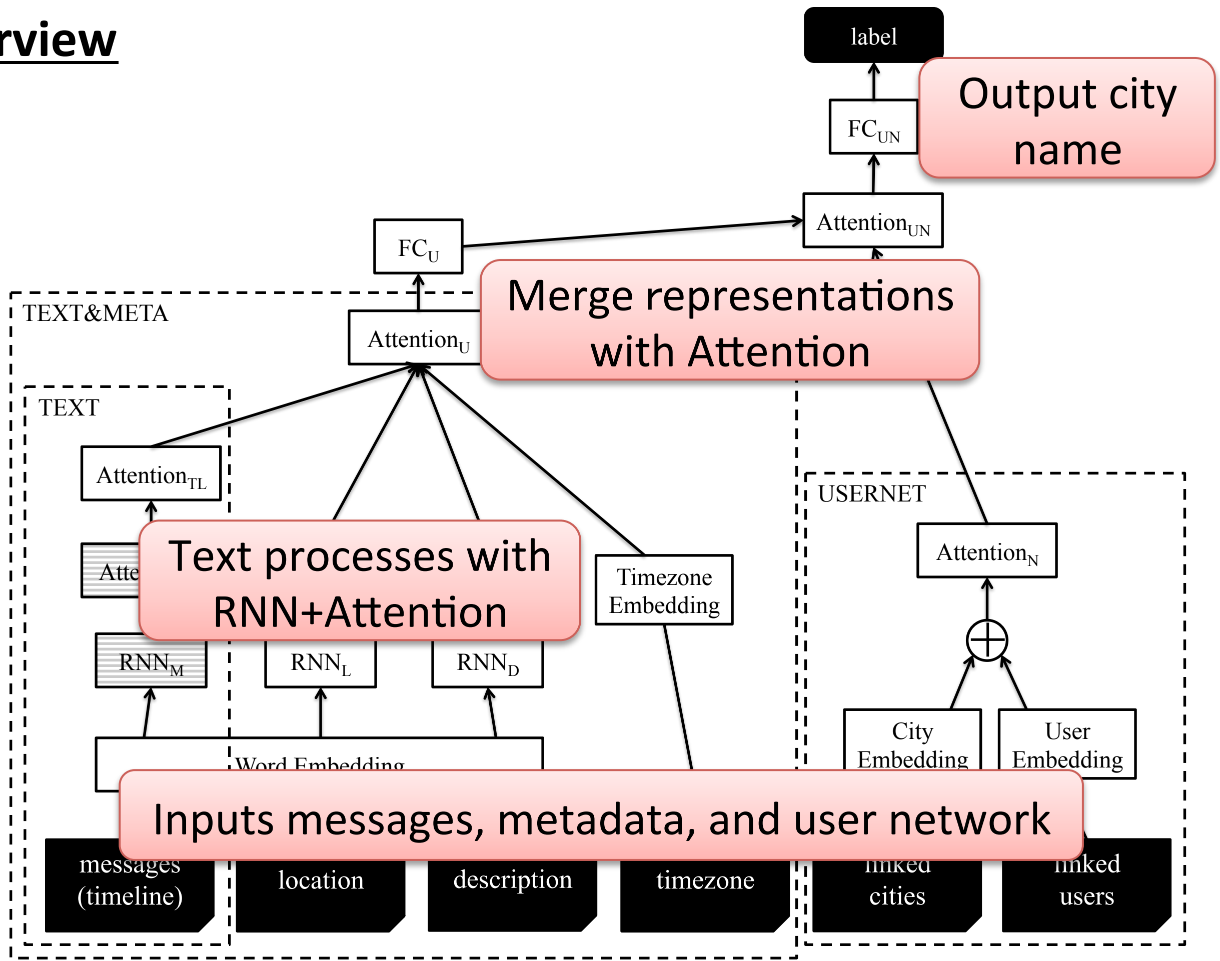
Overview



Our Approach

- Geolocation prediction with neural networks

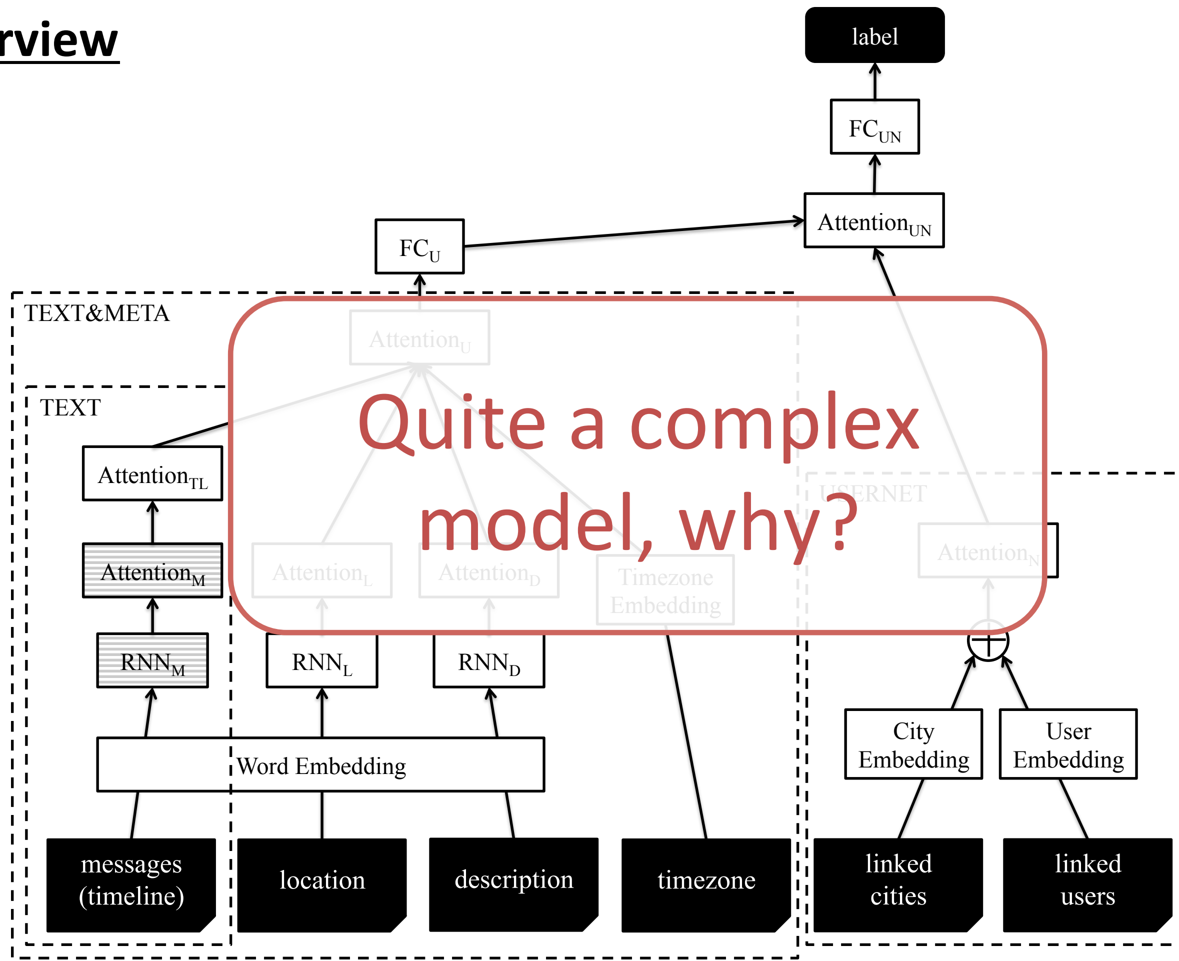
Overview



Our Approach

- Geolocation prediction with neural networks

Overview



Geolocation Prediction Target

- Twitter users
  - A popular target in previous works (Cheng et al., 2010; Eisenstein et al., 2010)

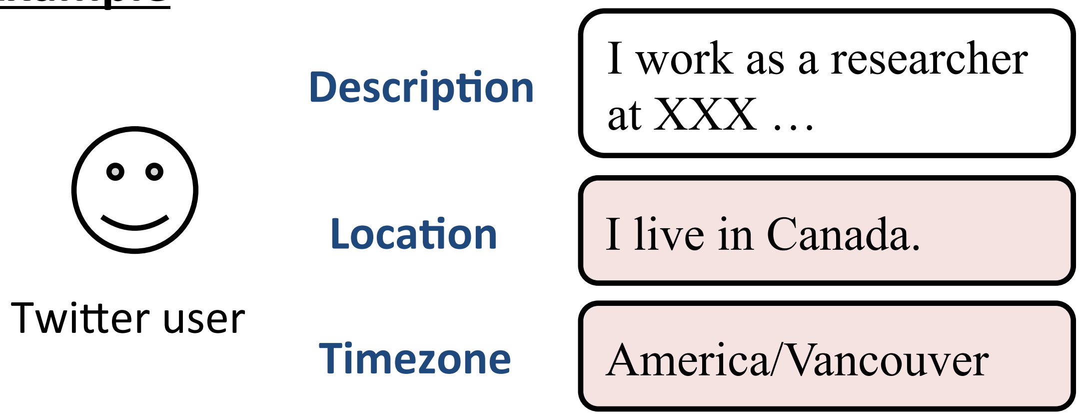
Geolocation Prediction Target

- Twitter users
  - A popular target in previous works (Cheng et al., 2010; Eisenstein et al., 2010)
- Characteristics
  - Multiple geolocation clues
    - Message (tweet)
    - Metadata
    - User network
  - Large-scale data
    - Ground-truth locations with geotags

Metadata

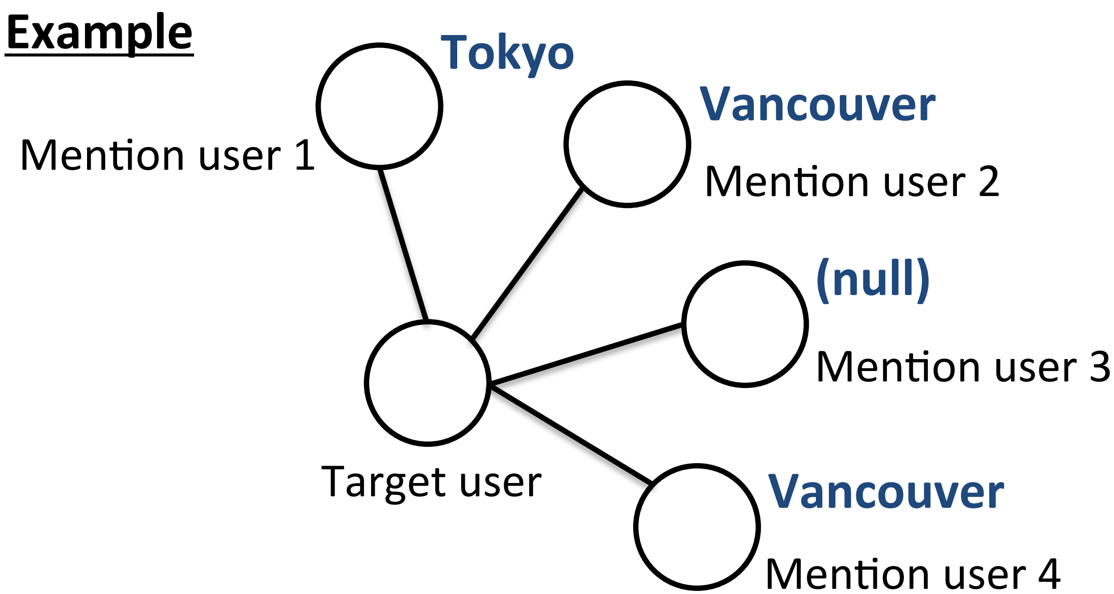
- Description, location, timezone, etc.
  - State-of-the-art performances combined with texts (Han et al., 2013, 2014; Jayasinghe et al., 2016; Miura et al., 2016)

Example



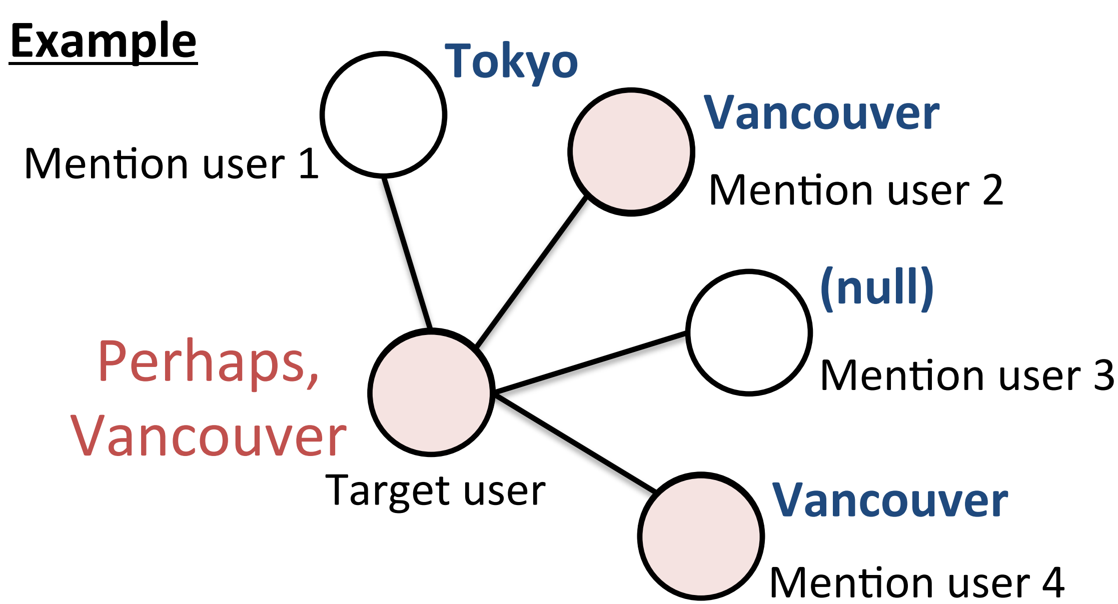
User Network

- Mention network, friend network, etc.
  - Prediction with label propagation
  - State-of-the-art performances combined with texts (Rahimi et al., 2015a; Jayasinghe et al., 2016)

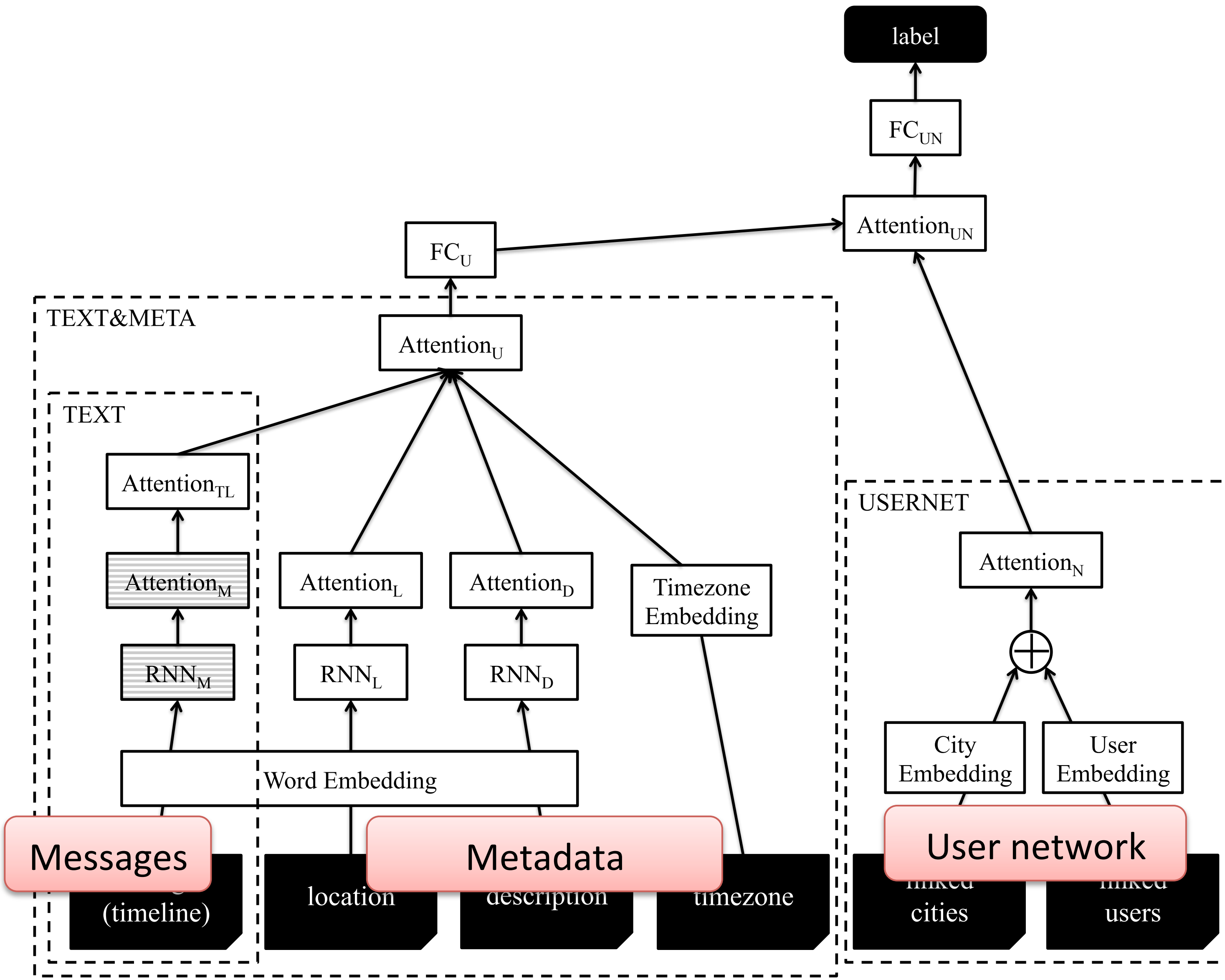


User Network

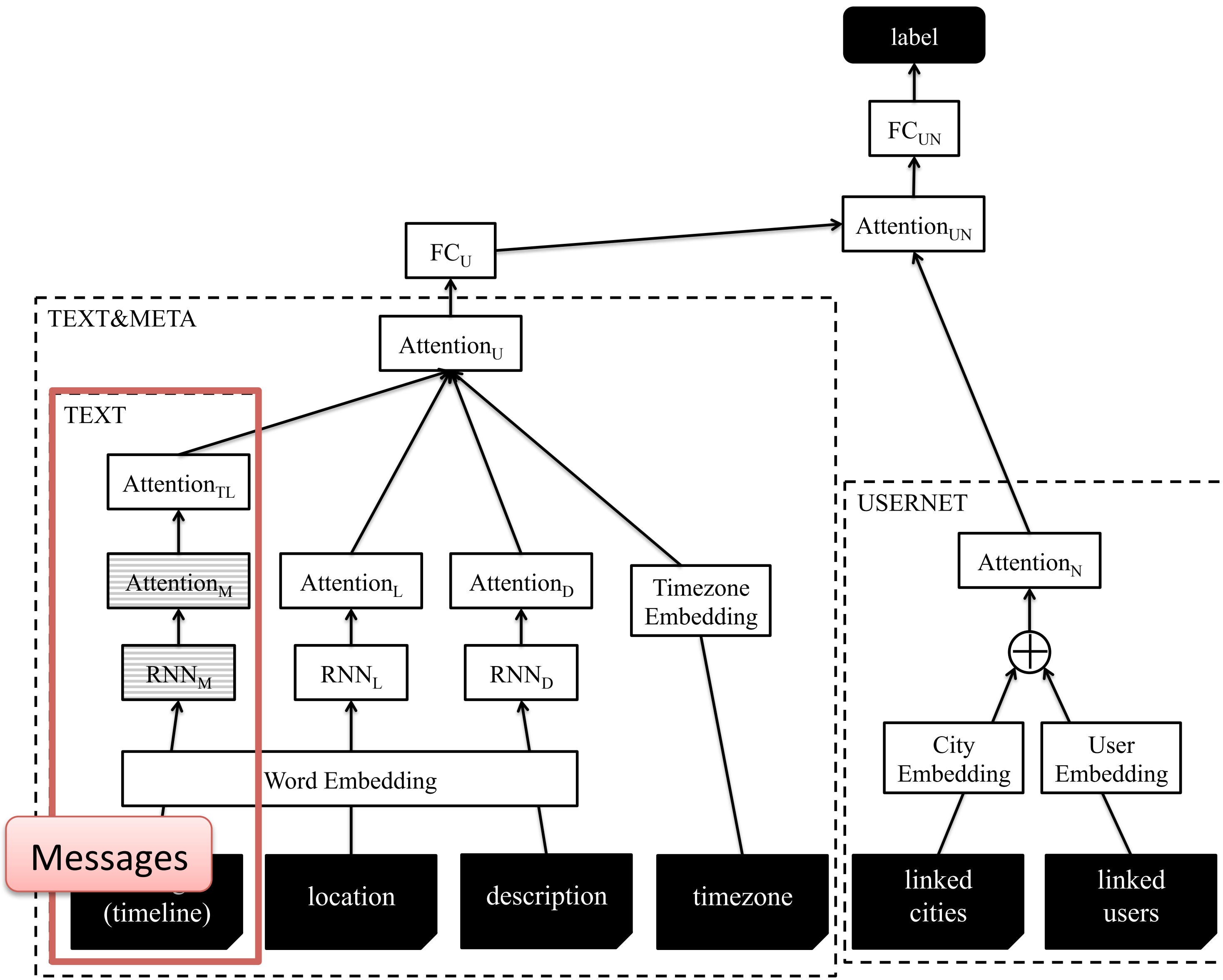
- Mention network, friend network, etc.
  - Prediction with label propagation
  - State-of-the-art performances combined with texts (Rahimi et al., 2015a; Jayasinghe et al., 2016)



Model


label
FC_{UN}
$Attention_{UN}$
FC_U
TEXT&META
$Attention_U$
TEXT
$Attention_{TL}$
$Attention_M$
RNN_M
$Attention_L$
RNN_L
$Attention_D$
RNN_D
Timezone Embedding
Word Embedding
(timeline)
location
description
timezone
USERNET
$Attention_N$
City Embedding
User Embedding
cities
users
Messages
Metadata
User network

TEXT Component (1)



TEXT Component (2)

- Hierarchical Neural Networks
  - Layers
    - Word Embedding
    - Recurrent Neural Network
      - Specifically, GRU (Cho et al., 2014)
    - (Self) Attention

Attention$_{\mathrm{M}}$

$$\boldsymbol{m} = \sum_t \alpha_t \boldsymbol{g}_t$$

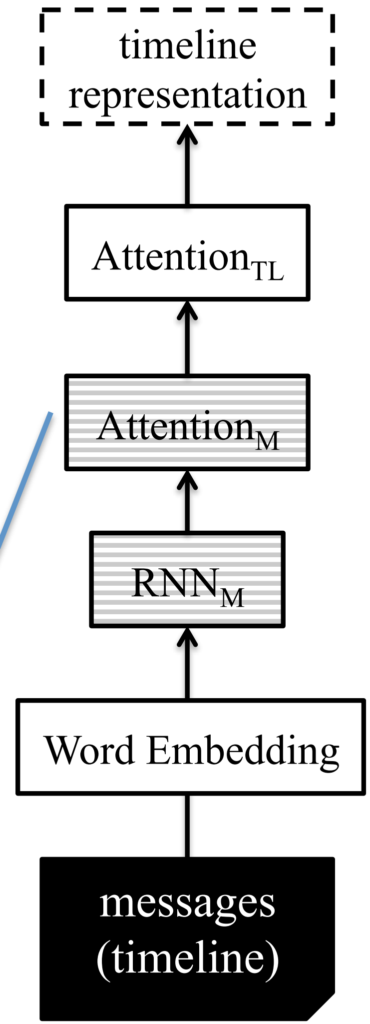
RNN$_{\mathrm{M}}$ output

Softmax

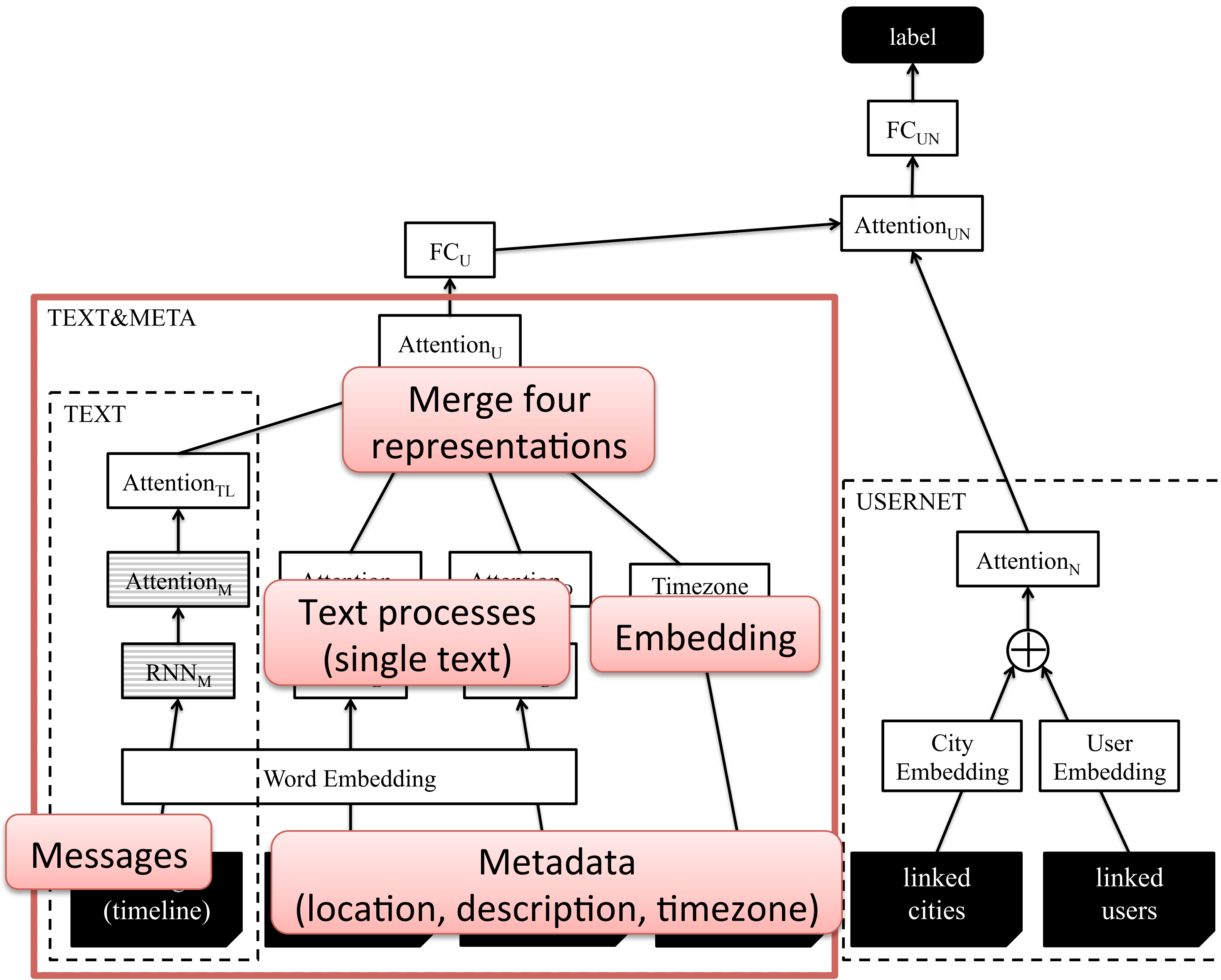
$$\alpha_t = \frac{\exp\left(\boldsymbol{v}_\alpha^T \boldsymbol{u}_t\right)}{\sum_t \exp\left(\boldsymbol{v}_\alpha^T \boldsymbol{u}_t\right)}$$

Multi-layer Perceptron

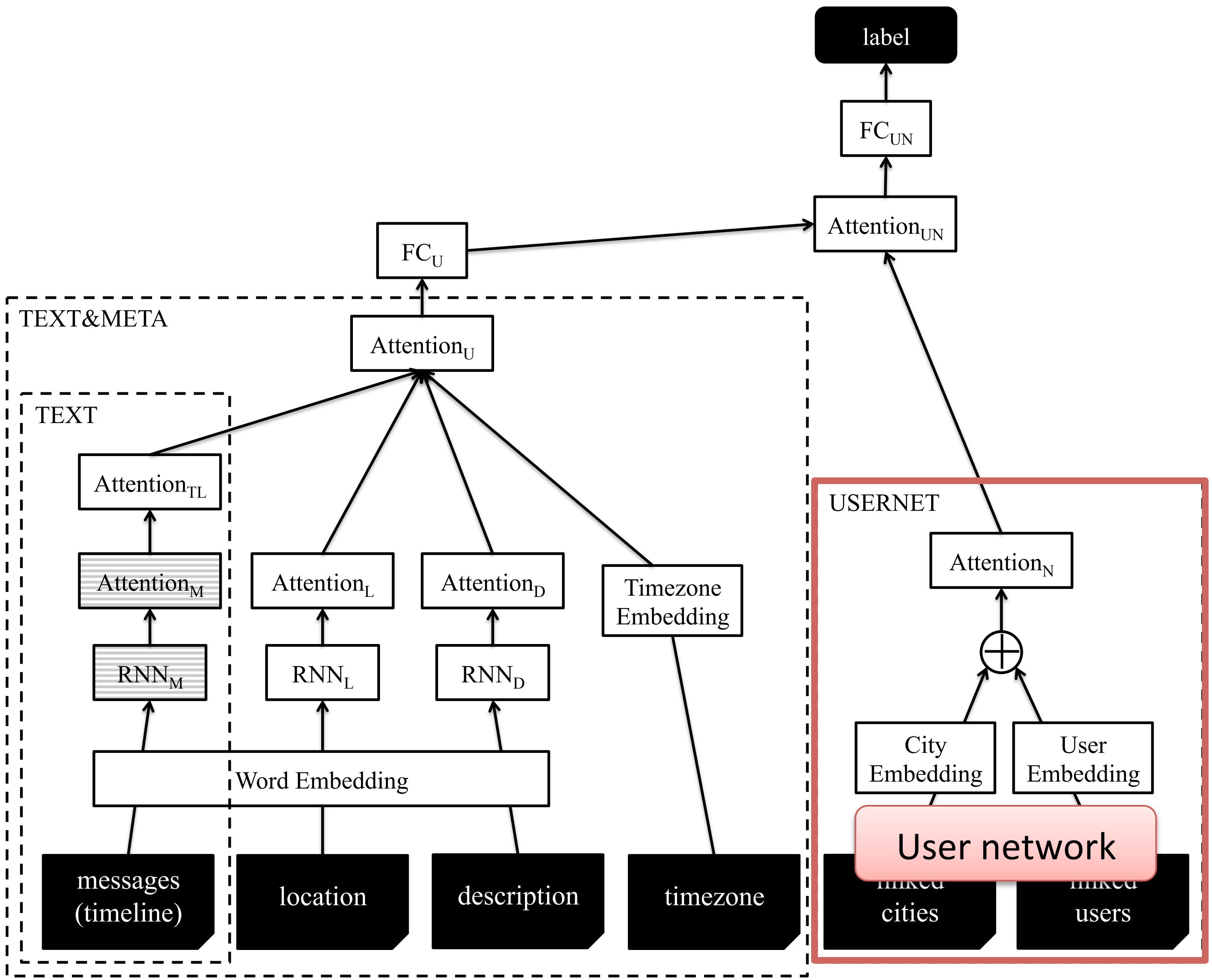
$$\boldsymbol{u}_t = \tanh\left(\boldsymbol{W}_\alpha \boldsymbol{g}_t + \boldsymbol{b}_\alpha\right)$$



TEXT&META Component

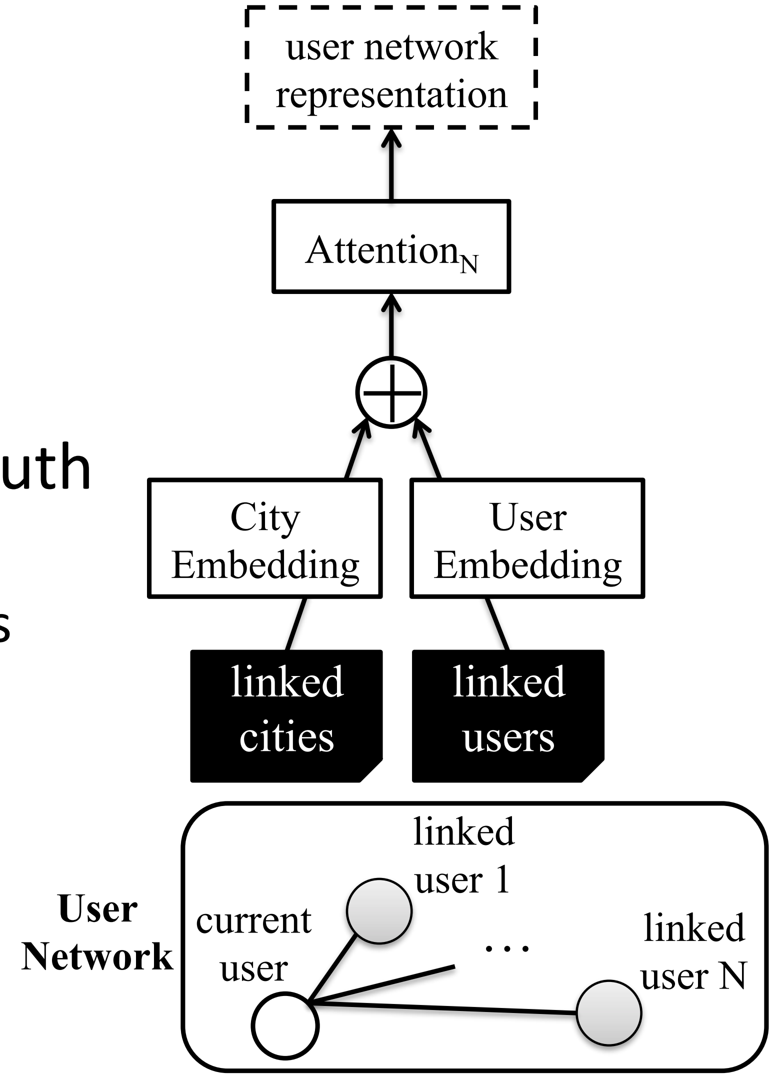


USERNET Component (1)

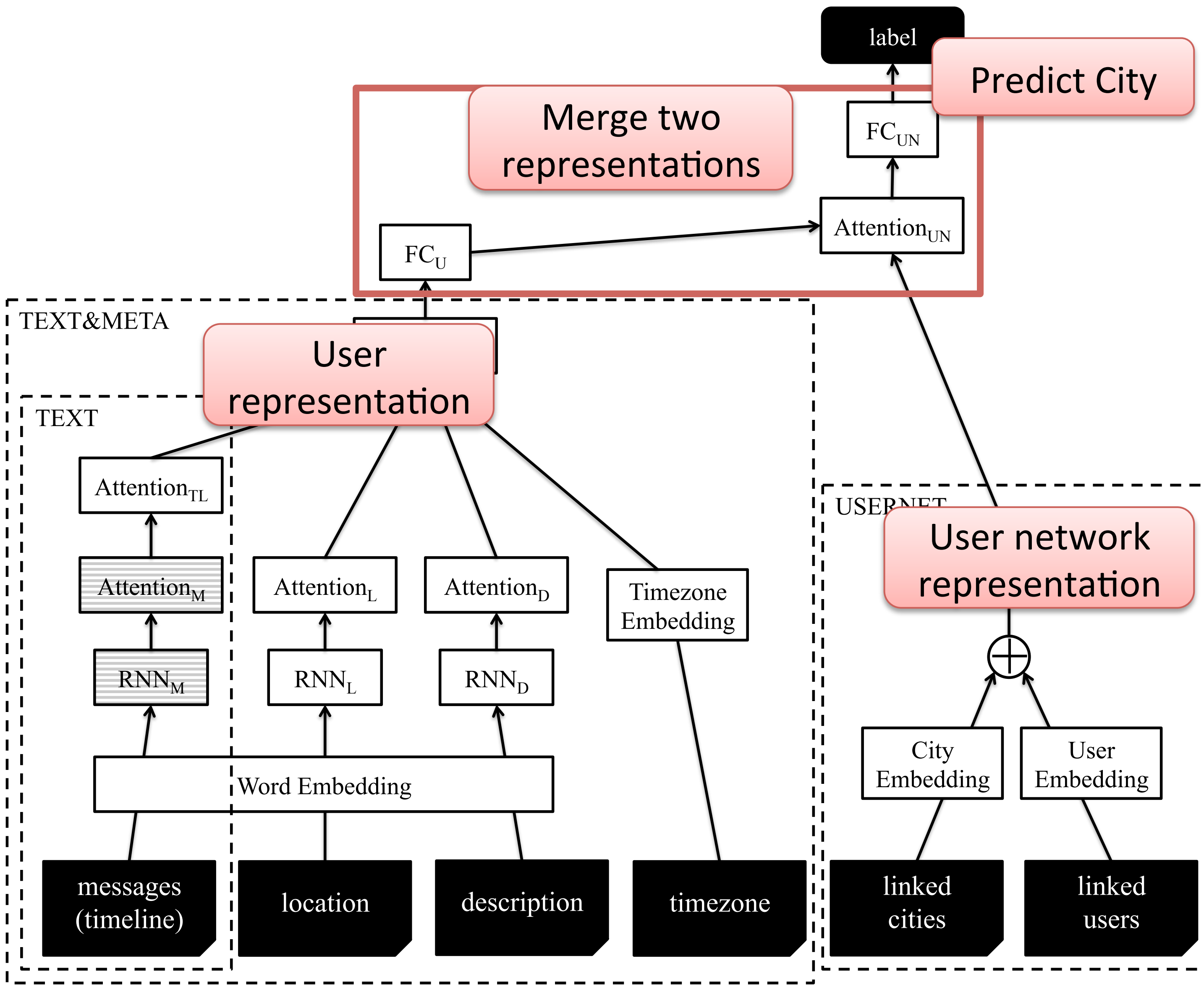


USERNET Component (2)

- Linked users
  - Embedding for each user
    - Regional celebrities
- Linked cities
  - Embedding for each ground-truth city of a user
    - “unknown” for unavailable cases
- Merge
  - Element-wise addition
  - Attention over N users



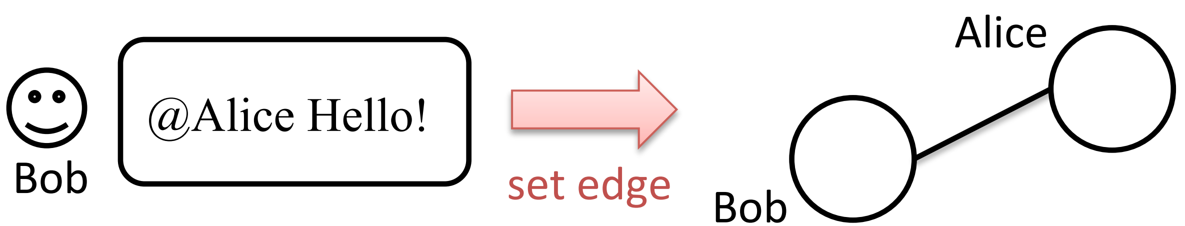
Unified Processes



Data (1)

- Two open datasets
  - TwitterUS (Roller et al., 2012)
  - W-NUT (Han et al., 2016)
- User Network
  - Uni-directional mention network (Rahimi et al., 2015a, 2015b)
    - Dataset users + one-hop users
    - Set undirected edges for mentions

Example



Data (2)

	TwitterUS (train)	W-NUT (train)
#user	279K	782K
#tweet	23.8M	9.03M
tweet/user	85.6	11.6
#reduced-edge*	2.11M	1.01M
reduced-edge/user	7.04	1.29
#city	339	3028

* Restricted edges to satisfy one of the following conditions:

1. Both users have ground truth locations.
2. One user has a ground truth location and another user is mentioned 5 times or more.

Baselines and Sub-models

Name	Text	Metadata	User Network	Key Components
LR	X			logistic regression, k-d tree (Rahimi et al. 2015a)
MADCEL-B-LR	X		X	logistic regression, k-d tree, Modified Adsorption (Rahimi et al. 2015a)
LR-STACK	X	X		logistic regression, k-d tree, stacking (Han et al. 2013, 2014)
MADCEL-B-LR-STACK	X	X	X	logistic regression, k-d tree, stacking, Modified Adsorption
SUB-NN-TEXT	X			TEXT
SUB-NN-UNET	X		X	TEXT, USERNET
SUB-NN-META	X	X		TEXT&META
Proposed Model	X	X	X	Full model

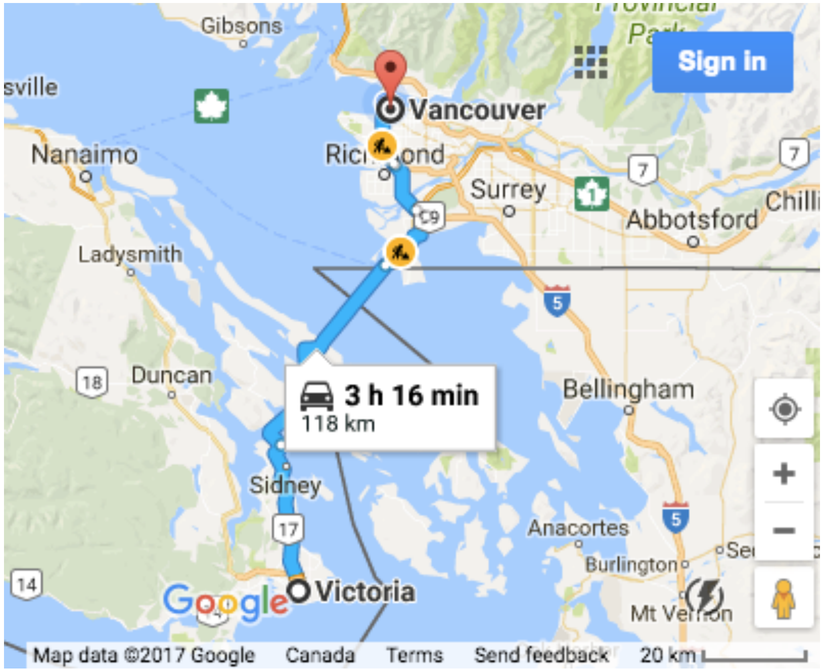
Metrics

Metric	Definition
Accuracy	The percentage of correctly predicted cities.
Accuracy@161	A relaxed accuracy that takes prediction errors within 161 km as correct predictions.
Median Error Distance*	Median value of error distances in predictions.
Mean Error Distance*	Mean value of error distances in predictions.

Example

Prediction: Vancouver
Ground-truth: Victoria

Accuracy = 0.0
Accuracy@161 = 1.0
Error Distance = 94km



* Error distance evaluations are excluded from this presentation for simplification.

Results (TwitterUS)

	Model	Accuracy	Accuracy @161
Baselines (implemented)	LR	42.0	52.7
	MADCEL-B-LR	50.2	60.1
	LR-STACK	50.8	64.1
	MADCEL-B-LR-STACK	55.7	67.7
Our Models	SUB-NN-TEXT	44.9**	55.6**
	SUB-NN-UNET	51.0	61.5*
	SUB-NN-META	54.6**	67.2**
	Proposed Model	58.5**	70.1**

+2.8% in accuracy

+2.4% in accuracy@161

* significant improvement with 5% significance level

** significant improvement with 1% significance level

Results (W-NUT)

	Model	Accuracy	Accuracy @161
Baselines (reported)	Jayasinghe et al. (2016)	52.6	-
Baselines (implemented)	LR	34.1	46.7
	MADCEL-B-LR	36.2	49.7
	LR-STACK	51.2	64.9
	MADCEL-B-LR-STACK	51.6	65.3
Our Models	SUB-NN-TEXT	35.4**	50.3**
	SUB-NN-UNET	38.1**	53.3**
	SUB-NN-META	54.7**	70.2**
	Proposed Model	56.4**	71.9**

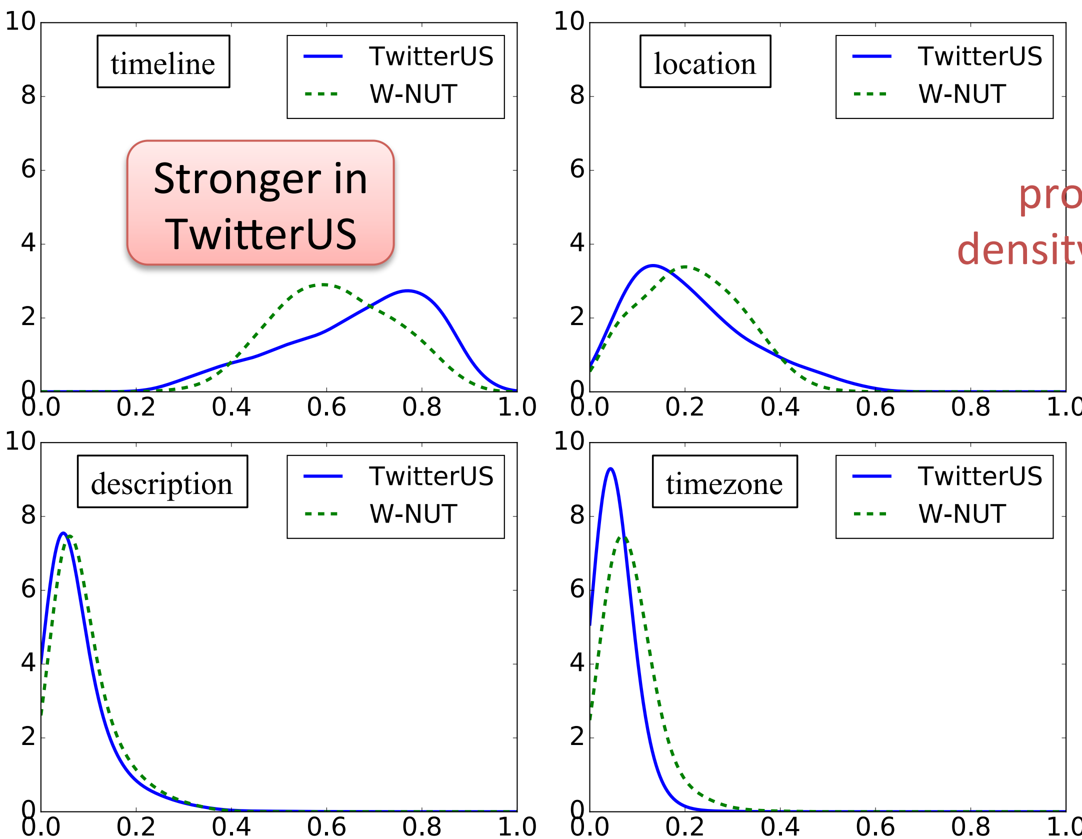
** significant improvement with 1% significance level

+3.8% in accuracy

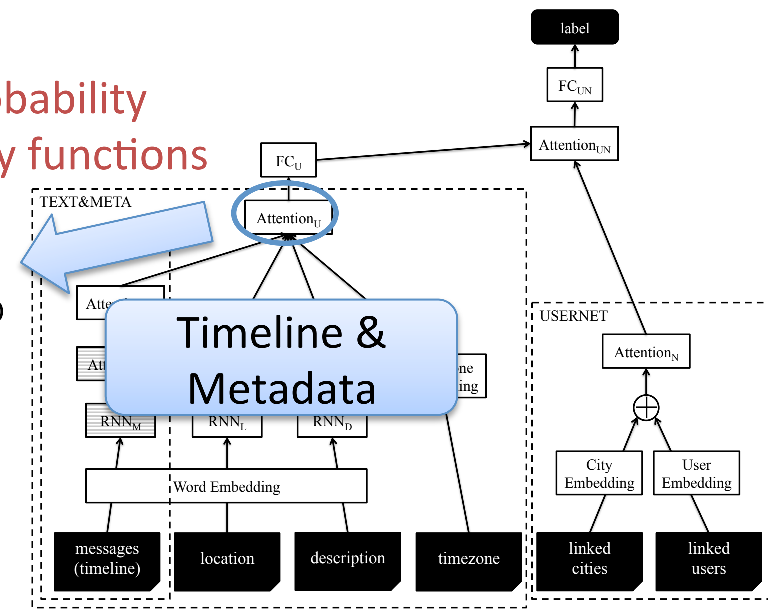
+4.8% in accuracy

+6.6% in accuracy@161

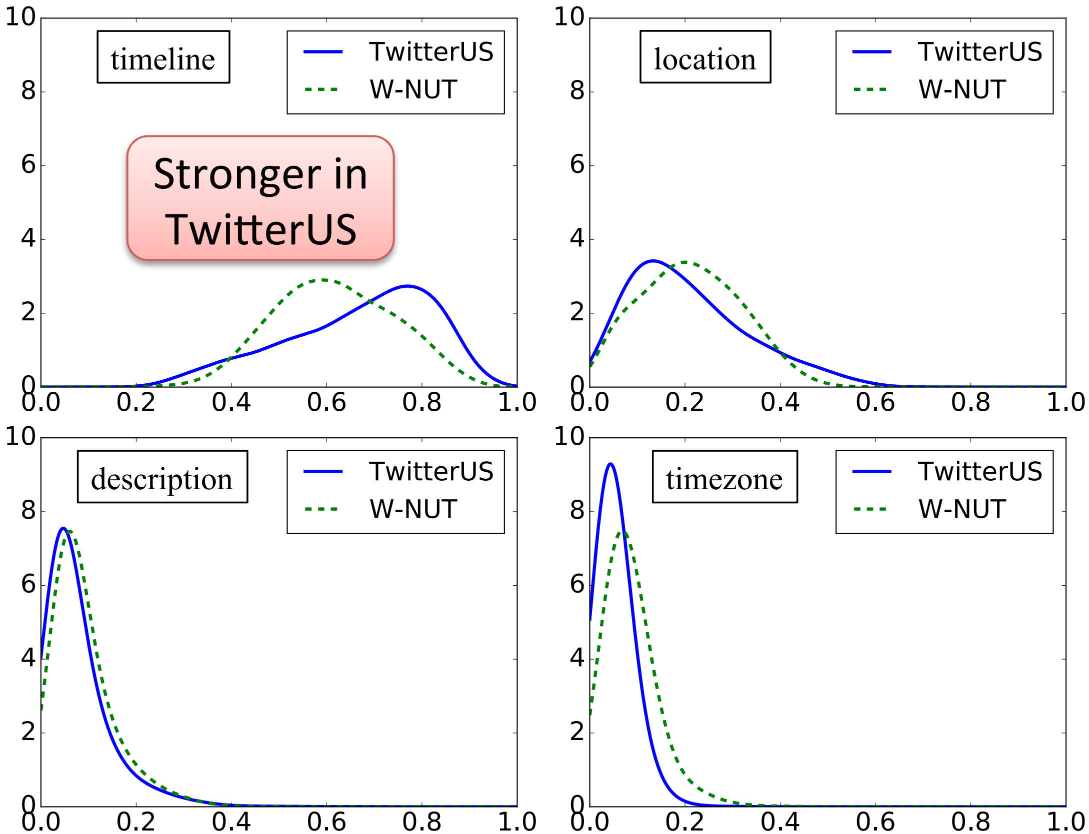
Analysis of Attention layers (1)



probability density functions

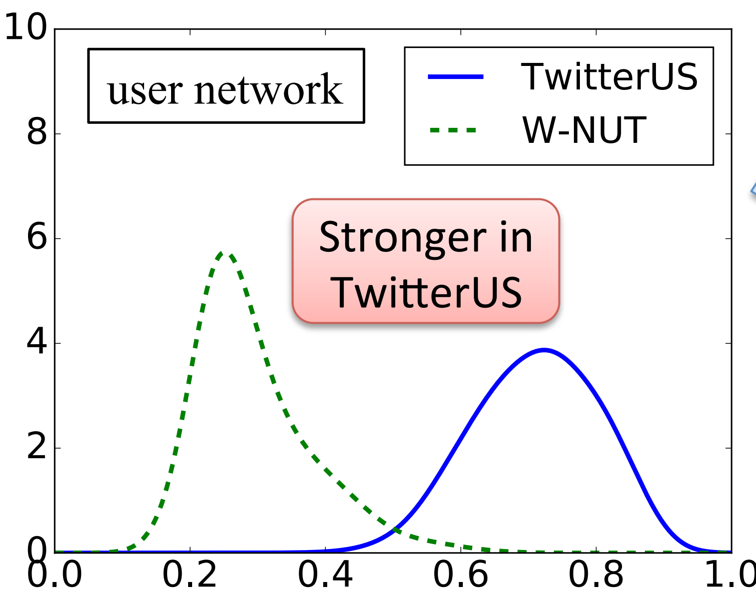


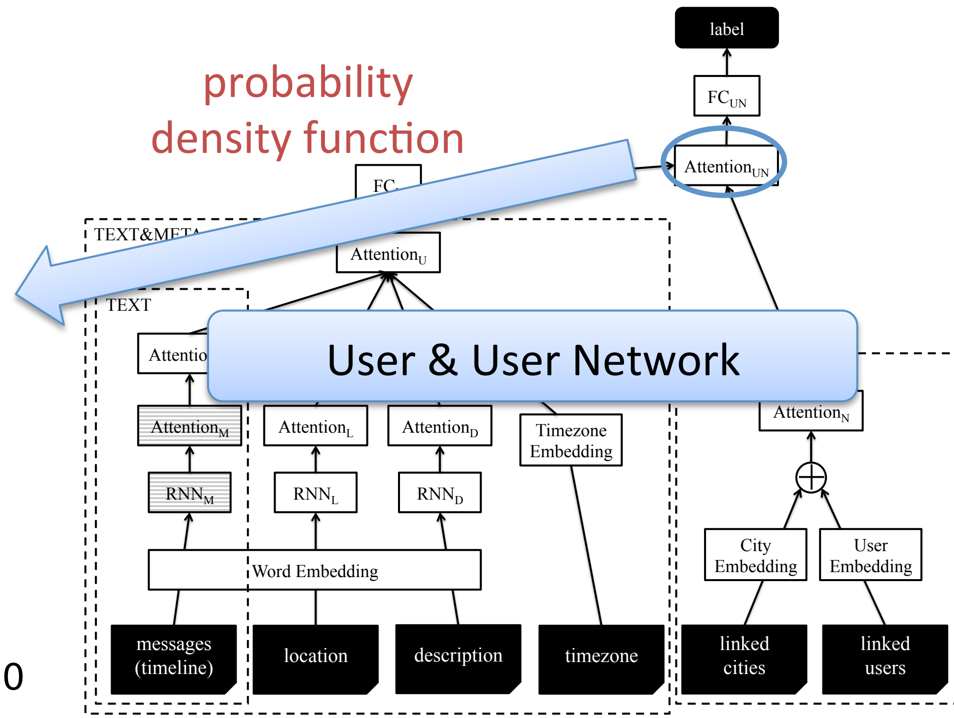
Analysis of Attention layers (1)



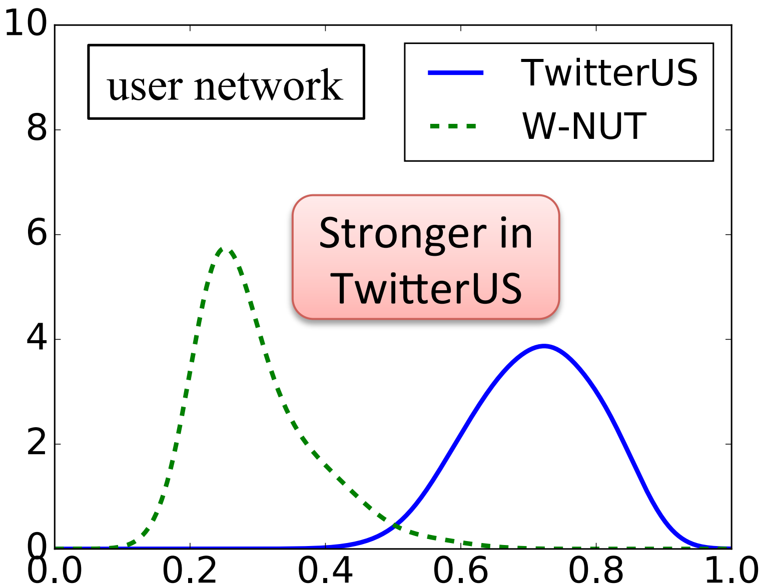
	TwitterUS (train)	W-NUT (train)
#user	279K	782K
#tweet	23.8M	9.03M
tweet/user	85.6	11.6

Analysis of Attention layers (2)





Analysis of Attention layers (2)



	TwitterUS (train)	W-NUT (train)
#reduced-edge	2.11M	1.01M
reduced-edge/user	7.04	1.29

Conclusion

- Proposed a neural network model for geolocation prediction
  - Unifies:
    - text
    - metadata
    - user network
  - Improvement over previous state-of-the art models
    - ✓ +2.8—3.8% in accuracy
    - ✓ +2.4—6.6% in accuracy@161
  - Analysis of attention probabilities:
    - Capturing statistical characteristics of the datasets

Future Works

- An extension of the proposed model
  - Introduction of temporal state
    - Capture location changes like travel
- Application to different tasks
  - For example, gender analysis or age analysis
    - Some metadata may not be effective

Thank You!

References (1)

Zhiyuan Cheng, James Caverlee, and Kyumin Lee. 2010. You are where you tweet: A content-based approach to geo-locating Twitter users. In Proc. of CIKM 2010. pp. 759–768.

Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Proc. of EMNLP 2014. pp. 1724–1734.

Jacob Eisenstein, Brendan O'Connor, Noah A. Smith, and Eric P. Xing. 2010. A latent variable model for geographic lexical variation. In Proc. of EMNLP 2010. pp.1277–1287.

Bo Han, Paul Cook, and Timothy Baldwin. 2013. A stacking-based approach to Twitter user geolocation prediction. In Proc. of ACL 2013: System Demonstrations. pp.7–12.

Bo Han, Paul Cook, and Timothy Baldwin. 2014. Text-based Twitter user geolocation prediction. Journal of Artificial Intelligence Research 49(1):451–500.

Bo Han, Afshin Rahimi, Leon Derczynski, and Timothy Baldwin. 2016. Twitter geolocation prediction shared task of the 2016 workshop on noisy user-generated text. In Proc. of W-NUT 2016. pp.213–217.

References (2)

Gaya Jayasinghe, Brian Jin, James Mchugh, Bella Robinson, and Stephen Wan. 2016. CSIRO Data61 at the WNUT geo shared task. In Proc. of W-NUT 2016. pp.218–226.

Yasuhide Miura, Motoki Taniguchi, Tomoki Taniguchi, and Tomoko Ohkuma. 2016. A simple scalable neural networks based model for geolocation prediction in Twitter. In Proc. of W-NUT 2016. pp.235–239.

Afshin Rahimi, Trevor Cohn, and Timothy Baldwin. 2015a. Twitter user geolocation using a unified text and network prediction model. In Proc. of ACL 2015. pp.630–636.

Afshin Rahimi, Duy Vu, Trevor Cohn, and Timothy Baldwin. 2015b. Exploiting text and network context for geolocation of social media users. In Proc. of NAACL-HLT 2015. pp.1362–1367.

Stephen Roller, Michael Speriosu, Sarat Rallapalli, Benjamin Wing, and Jason Baldridge. 2012. Supervised text-based geolocation using language models on an adaptive grid. In Proc. of EMNLP 2012. pp.1500–1510.

Any Questions?

Motivation

- Crucial for tasks like:
  - Disaster Analysis
  - Disease Analysis
  - Political Analysis
- Enables a region specific analysis in:
  - Sentiment Analysis
  - User Attribute Analysis

Other Geolocation Targets

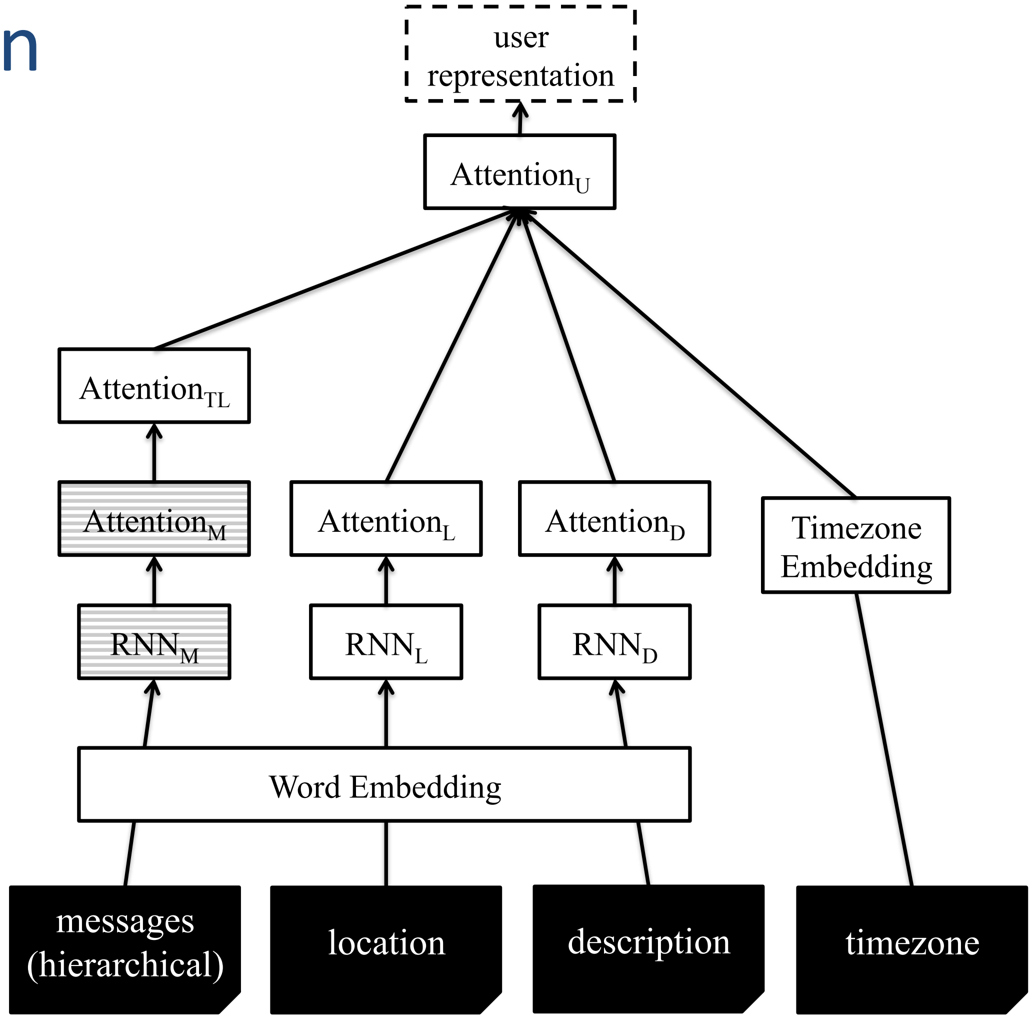
- In previous works:
  - Wikipedia articles (Overell, 2009)
  - Flicker photos (Serdyukov et al., 2009; Crandall et al., 2009)
  - Facebook users (Backstrom et al., 2010)
  - Twitter users (Cheng et al., 2010; Eisenstein et al., 2010)

Comparison with Previous Approaches

- In previous works:
  - Ensemble approaches
    - Stacking (Han et al., 2013, 2014)
    - Dongle nodes (Rahimi et al., 2015a, 2015b)
    - Cascades ensemble (Jayasinghe et al., 2016)
- This work:
  - Neural network
    - Multiple inputs
    - Attention mechanism to merge inputs

Text and Metadata Component (2)

- Location, Description
  - RNN+Attention
    - Single text
- Timezone
  - Embedding for each timezone
- Merge
  - Attention over four representations



Data (1)(detailed)

- Two open datasets
  - TwitterUS
    - The dataset of Roller et al. (2012)
    - 449K Twitter users
    - North America region
  - W-NUT
    - The dataset of W-NUT 2016 geolocation prediction shared task (Han et al., 2016)
    - 1.02M Twitter users
    - World-wide

Data (2)(detailed)

	TwitterUS (train)	W-NUT (train)
#user	279K	782K
#tweet	23.8M	9.03M
tweet/user	85.6	11.6
#edge	3.69M	3.21M
#reduced-edge*	2.11M	1.01M
reduced-edge/user	7.04	1.29
#city	339	3028

* Restricted edges to satisfy one of the following conditions:

1. Both users have ground truth locations.
2. One user has a ground truth location and another user is mentioned 5 times or more.

Baselines

- LR
  - l_1-regularized logistic regression with k-d tree regions (Roller et al., 2012; Rahimi et al. 2015a)
- MADCEL-B-LR
  - LR with Modified Adsorption (Rahimi et al. 2015a)
    - Modified Adsorption is a graph-based label propagation algorithm
- LR-STACK
  - Stacking of logistic regression classifiers
    - Four LR classifier for messages and metadata
    - l_2-regularized logistic regression meta-classifier
  - Similar to previous stacking approaches (Han et al., 2013, 2014)
- MADCEL-B-LR-STACK
  - LR-STACK with Modified Adsorption

Results (TwitterUS)(2)

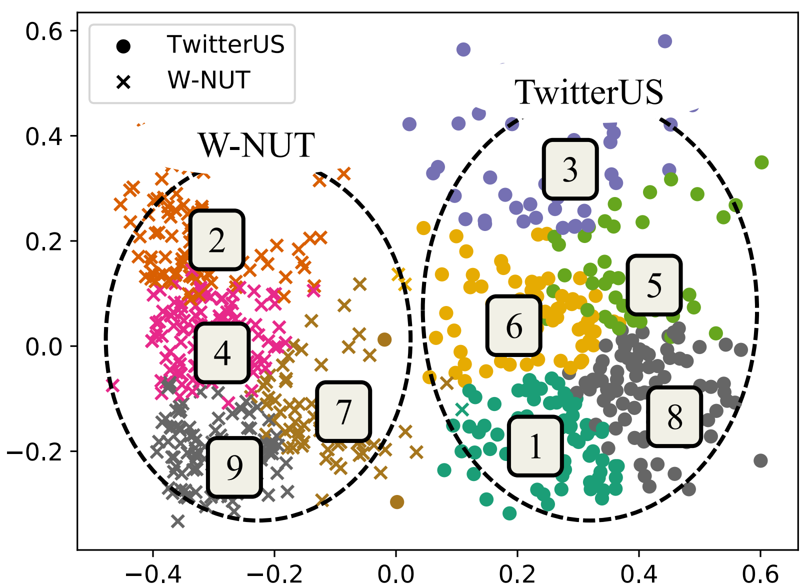
	Model	Sign. Test ID	Accuracy	Accuracy @161	Error Distance Median	Error Distance Mean
Baselines (reported)	Han et al. (2012)		26.0	45.0	260	814
	Wing and Baldridge (2014)		-	49.2	170.5	703.6
	LR (Rahimi et al. 2015b)		-	50	159	686
	LR-NA (Rahimi et al. 2016)		-	51	148	636
	MADCEL-B-LR (Rahimi et al. 2015a)		-	60	77	533
	MADCEL-W-LR (Rahimi et al. 2015a)		-	60	78	529
Baselines (implemented)	LR	i	42.0	52.7	121.1	666.6
	MADCEL-B-LR	ii	50.2	60.1	66.5	582.8
	LR-STACK	iii	50.8	64.1	42.3^{*}	427.7
	MADCEL-B-LR-STACK	iv	55.7	67.7	45.1	412.7
Our Models	SUB-NN-TEXT	i	44.9^{**}	55.6^{**}	110.5	585.1^{**}
	SUB-NN-UNET	ii	51.0	61.5^{*}	65.0	481.5^{**}
	SUB-NN-META	iii	54.6^{**}	67.2^{**}	46.8	356.3^{**}
	Proposed Model	iv	58.5^{**}	70.1^{**}	41.9^{*}	335.7^{**}

Results (W-NUT)(2)

	Model	Sign. Test ID	Accuracy	Accuracy @161	Error Distance	
					Median	Mean
Baselines (reported)	Miura et al. (2016)		47.6	-	16.1	1122.3
	Jayasinghe et al. (2016)		52.6	-	21.7	1928.8
Baselines (implemented)	LR	i	34.1	46.7	248.7	2216.4
	MADCEL-B-LR	ii	36.2	49.7	166.3	2120.6
	LR-STACK	iii	51.2	64.9	0.0	1496.4
	MADCEL-B-LR-STACK	iv	51.6	65.3	0.0	1471.9
Our Models	SUB-NN-TEXT	i	35.4**	50.3**	155.8**	1592.6**
	SUB-NN-UNET	ii	38.1**	53.3**	99.9**	1498.6**
	SUB-NN-META	iii	54.7**	70.2**	0.0	825.8**
	Proposed Model	iv	56.4**	71.9**	0.0	780.5**

Attention Patterns

- Clustering of Attention_{U} and Attention_{UN}
  - k-means with 9 clusters
  - Some typical patterns
    - Cluster 2 and Cluster 3 balancing Timeline and Location



Cluster ID	Dataset	Time-line	Loca-tion	Descr-iption	Time-zone	User	User Net
1	TwitterUS	0.843	0.082	0.040	0.035	0.359	0.641
2	W-NUT	0.517	0.317	0.081	0.085	0.732	0.268
3	TwitterUS	0.432	0.430	0.069	0.069	0.319	0.681
4	W-NUT	0.637	0.160	0.097	0.105	0.737	0.263
5	TwitterUS	0.593	0.219	0.114	0.075	0.230	0.770
6	TwitterUS	0.672	0.214	0.069	0.045	0.365	0.635
7	W-NUT	0.741	0.077	0.080	0.102	0.605	0.395
8	TwitterUS	0.766	0.099	0.068	0.067	0.222	0.778
9	W-NUT	0.800	0.067	0.056	0.078	0.730	0.270

Errors with High Confidences

- Incorrect Location Field
  - Ex.
    - Location Field: Hong Kong
    - Ground-truth: Tronto
  - Perhaps, a house move
- Travel
  - Ex.
    - Tweeting about “San Francisco”
    - Ground-truth: Boston

Model Configuration (1)

	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	-

Model Configuration (2)

- Pre-training
  - Word Embeddings
    - skip-gram
      - learning rate=0.025, window size=5, negative sample size=5, epoch=5
  - Network Embeddings
    - LINE
      - initial learning rate=0.025, order=2, negative sample size=5, training sample size=100M
- Optimization
  - Adam
  - l_2 regularization

Training Time

- GPU
  - GeForce GTX Titan X
  - Titan X
- Time
  - 15—20 hours
    - Full model
    - Longer for TwitterUS