AAAI 2019

Cooperative Multimodal Approach to Depression Detection in Twitter

Tao Gui*, Liang Zhu*
Qi Zhang, Minlong Peng, Xu Zhou, Keyu Ding, Zhigang Chen

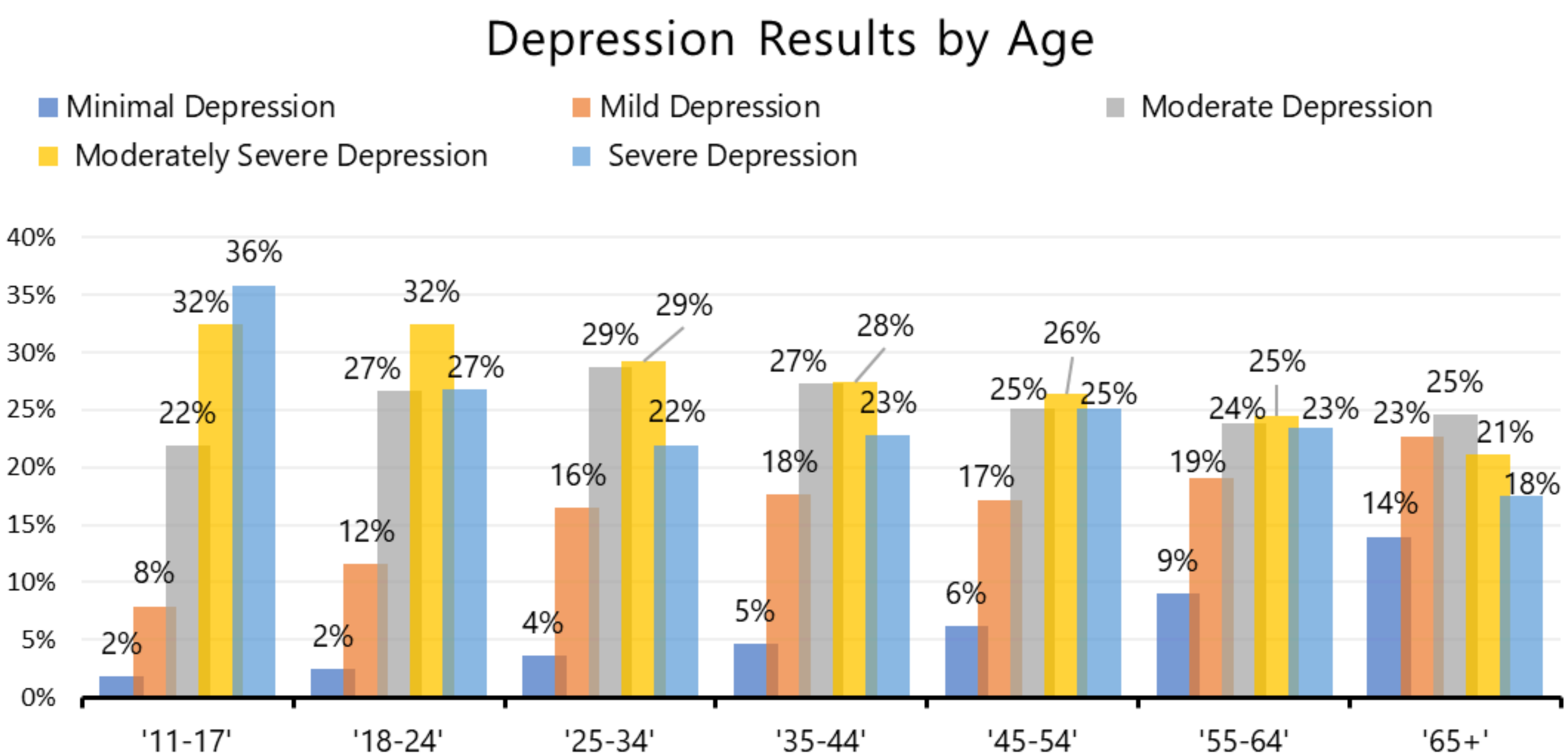
*Equal contribution

Outline

1. Introduction and Motivation
2. Approach
3. Experiments
4. Conclusion

Introduction and Motivation

Background

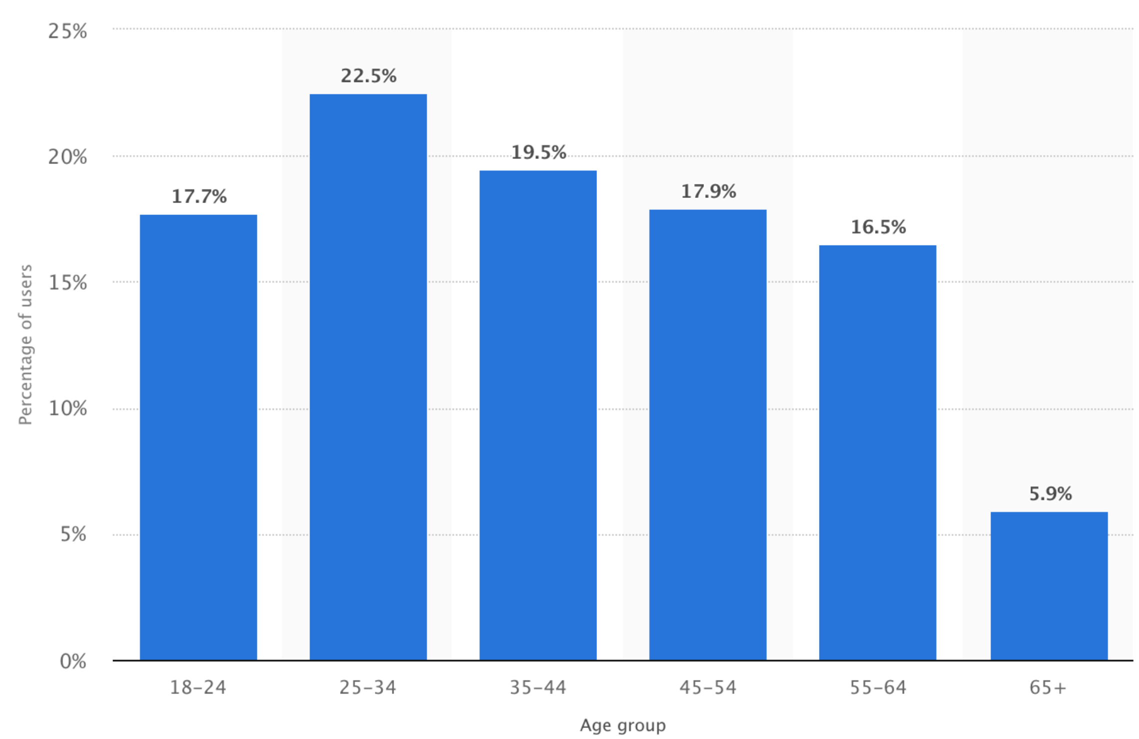


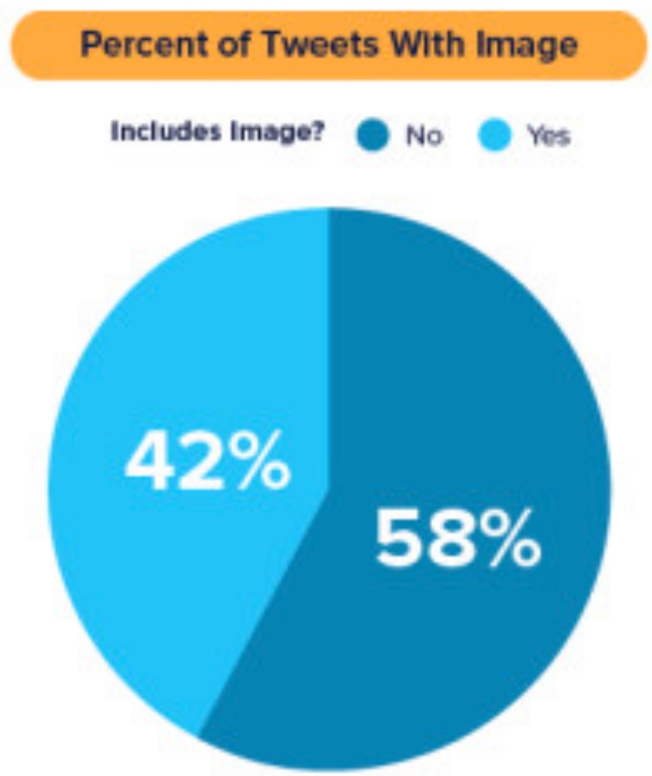
- Nowadays, depression is common.
- Our youth is at greatest risk of depression and self-harm.

http://www.mentalhealthamerica.net/online-depression-screening-data-and-statistics

Introduction and Motivation

Background



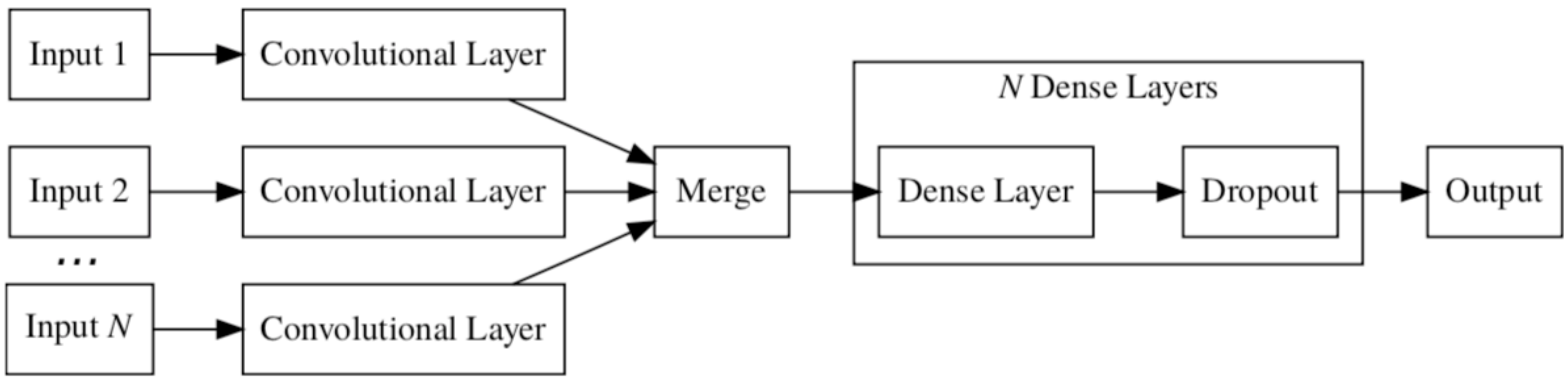


- A mount of Tweets has images.
- Young people are willing to post on Twitter.

https://www.statista.com/statistics/192703/age-distribution-of-users-on-twitter-in-the-united-states/
https://thenextweb.com/socialmedia/2015/11/03/what-analyzing-1-million-tweets-taught-us/

Introduction and Motivation

Related Work

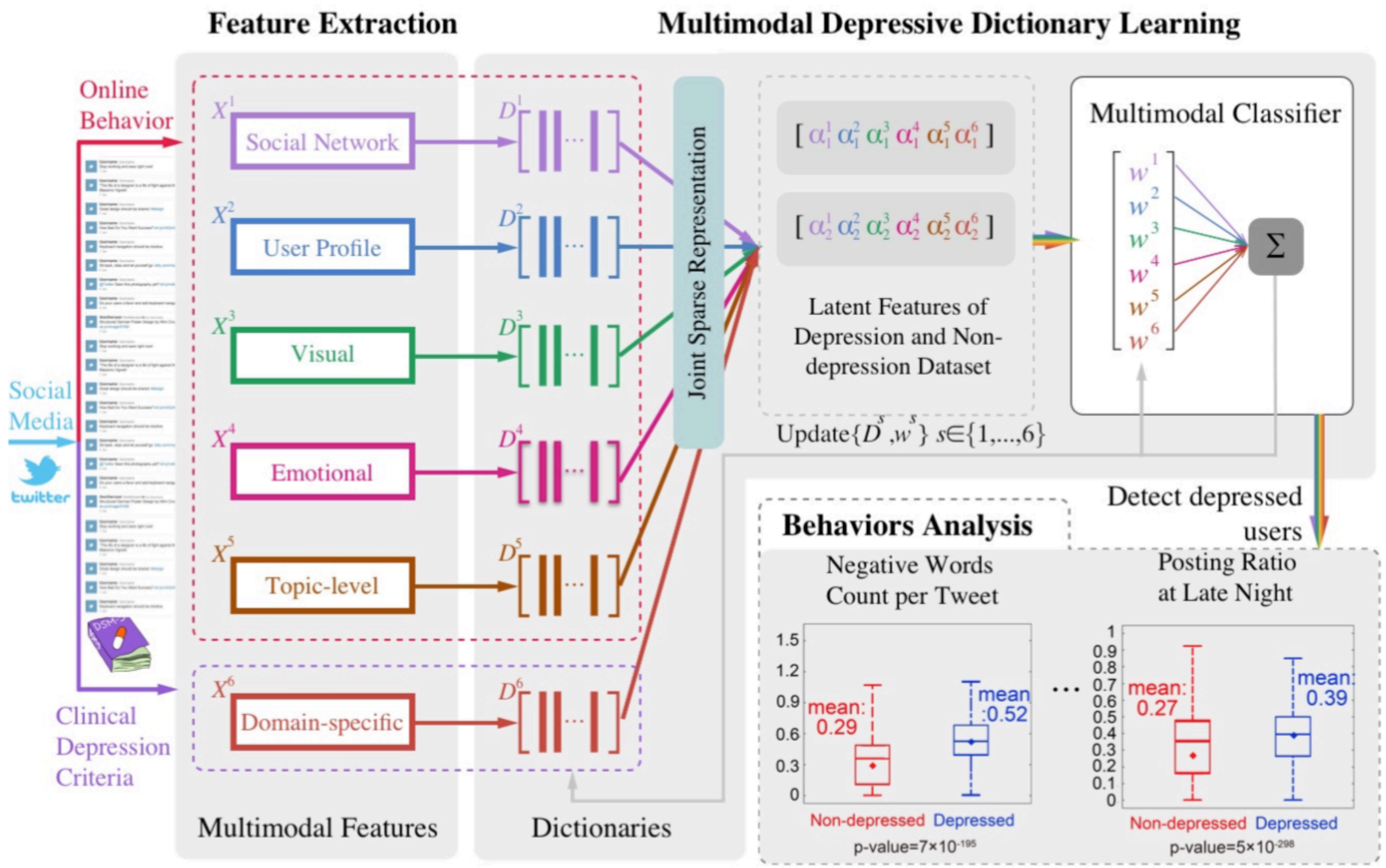


Depression and Self-Harm Risk Assessment in Online Forums

Yates, Cohan, and Goharian (EMNLP 2017)

Introduction and Motivation

Related Work



Depression detection via harvesting social media: A multimodal dictionary learning solution.

Shen et al. (IJCAI 2017)

Introduction and Motivation

Motivation

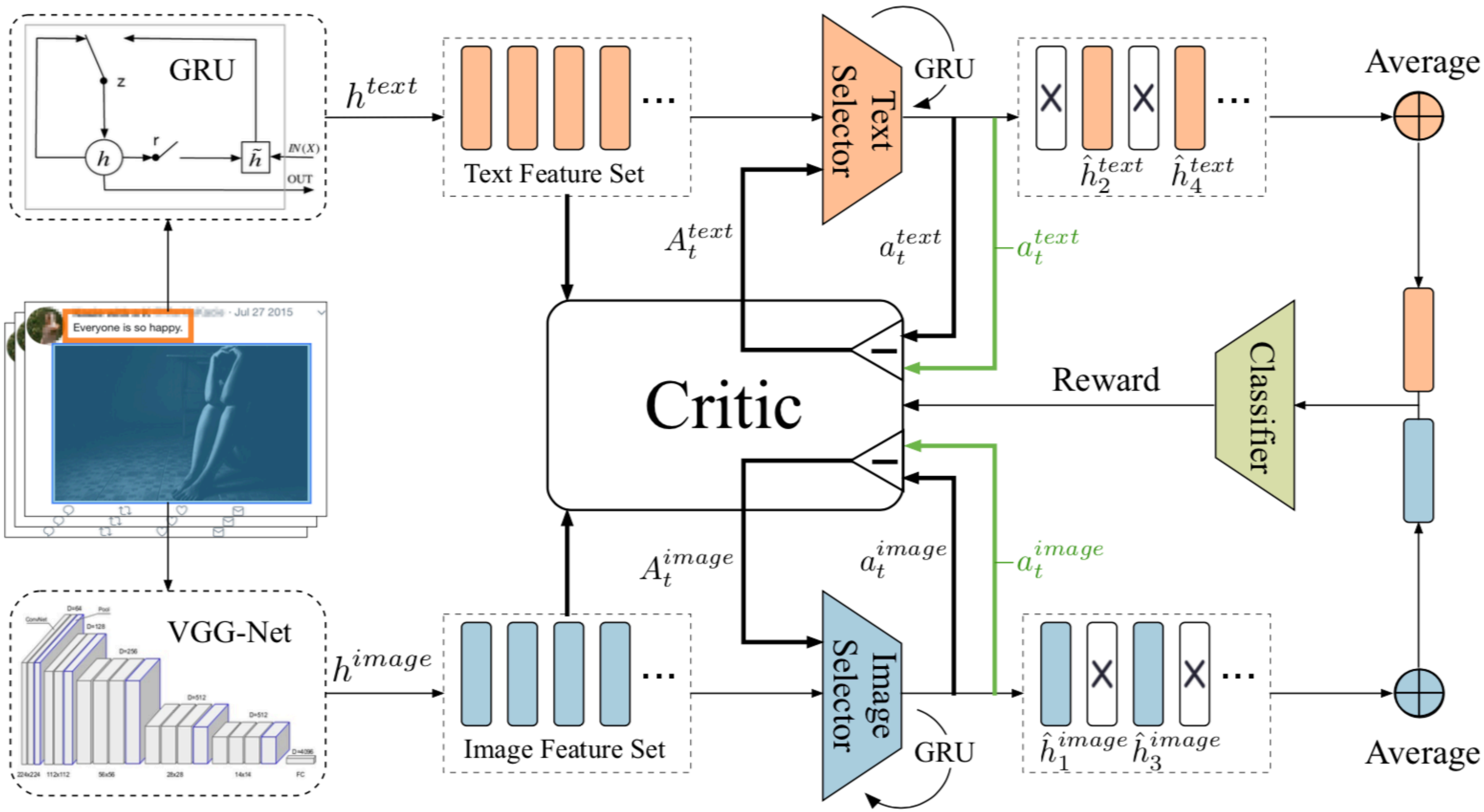
- Jointly consider textual and visual information.
- Extract relevant indicator texts and images.

Outline

1. Introduction and Motivation
2. Approach
3. Experiments
4. Conclusion

Approach

Our model



Our proposed model: COMMA

Approach

Approach

Algorithm 1 COMMA for Depression Detection

1: Randomly initialize critic network $Q(S, \pi, \mathbf{a}|\theta_Q)$ and two selectors $\pi(s|\theta^e_\pi)$ with weights θ_Q and θ^e_π.
2: Initialize target network Q' and π' with weights $\theta_{Q'} \leftarrow \theta_Q$, $\theta^e_{\pi'} \leftarrow \theta^e_\pi$. Initialize replay buffer R
3: **for** episode = 1, M **do**
4: Receive initial observation state h^e_1
5: **for** $t = 1, T$ **do**
6: Select action $a^e_t = \pi(h^e_t|\theta^e_\pi)$ according to the current policy
7: Execute action a^e_t and observe the likelihood of ground truth $\Pr(y = \hat{y}_u|o_t)$ and observe the new state h^e_{t+1}
8: Execute action a^e_{t+1} and observe the likelihood of ground truth $\Pr(y = \hat{y}_u|o_{t+1})$, thereby obtain the reward $r_t = \Pr(y = \hat{y}_u|o_{t+1}) - \Pr(y = \hat{y}_u|o_t)$
9: Store transition $(H^t_{init}, A_t, r_t, H^{t+1}_{init})$ in R

Approach

Approach

Sample a random minibatch of N transitions $(H^i_{init}, A_i, r_i, H^{i+1}_{init})$ from R

Set $z_i = r_i + \gamma Q'(H^{i+1}_{init}, \Pi_{i+1}, A_{i+1})$

Update critic by minimizing the loss: $\mathcal{L}(\theta_Q) = \frac{1}{N}\sum_i [z_i - Q(H^i_{init}, \Pi_i, A_i|\theta_Q)]^2$

Update selectors using differentiated advantages:

$$A^e(H, \Pi, A) = Q(H, \Pi, A) - Q(H, \Pi, (-a^e, a^{-e}))$$

$$\nabla_{\theta^e_\pi} J(\theta^e_\pi) = \nabla_{\theta^e_\pi} \log \pi(a^e_t | h^e_t) A^e(H, \Pi, A)$$

Update the target networks:

$$\theta_{Q'} = \tau\theta_Q + (1-\tau)\theta_{Q'}, \theta^e_{\pi'} = \tau\theta^e_\pi + (1-\tau)\theta^e_{\pi'}$$

end for

Update the depression classifier by minimizing the cross entropy loss:

$$J(\theta_C) = -[y_u \log \hat{y}_u + (1-y_u)\log(1-\hat{y}_u)]$$

end for

Outline

1. Introduction and Motivation
2. Approach
3. Experiments
4. Conclusion

Experiments

Dataset

Dataset		# Users	# T	# T + I
D_1	Depressed	1,402	292,564	-
	Non-Depressed	5,160	3,953,183	-
D_2	Depressed	1,402	251,834	40,730
	Non-Depressed	5, 160	3,302,366	650,817

Table 1: Statistical details of the datasets used in our experiments, where **# T** and **# T + I** represent the number of tweets that contain only texts and that contain both text + image pairs, respectively.

Based on the dataset provided by Shen et al. (IJCAI 2017), we constructed a dataset containing both tweets and images.

Experiments

Results

Methods	Training Data	Accuracy	Precision	Recall	F1
NB (Pedregosa et al. 2011)	Various Features	0.724	0.727	0.728	0.728
MSNL (Song et al. 2015)		0.818	0.818	0.818	0.818
WDL (Rolet, Cuturi, and Peyré 2016)		0.768	0.769	0.768	0.768
MDL (Shen et al. 2017)		0.848	0.848	0.850	0.849
GRU (Chung et al. 2014)	Text	0.824	0.825	0.823	0.824
GRU + Random sampling		0.760	0.760	0.757	0.756
VGG-Net (Simonyan and Zisserman 2014)	Image	0.702	0.703	0.702	0.702
VGG-Net + Random sampling		0.642	0.643	0.642	0.643
GRU + VGG-Net	Text+Image	0.845	0.843	0.847	0.845
GRU + VGG-Net + Random sampling		0.811	0.811	0.810	0.810
Co-Attention (Lu et al. 2016)		0.866	0.871	0.863	0.865
Dual-Attention (Nam, Ha, and Kim 2017)		0.848	0.848	0.848	0.848
Modality Attention (Moon, Neves, and Carvalho 2018)		0.866	0.868	0.862	0.864
GRU + VGG-Net + Unified advantages (Egorov 2016)		0.866	0.866	0.865	0.865
GRU + VGG-Net + COMMA (text + image)		**0.900**	**0.900**	**0.901**	**0.900**

Table 2: Comparison of performances in terms of four selected measures.

Experiments

Analysis

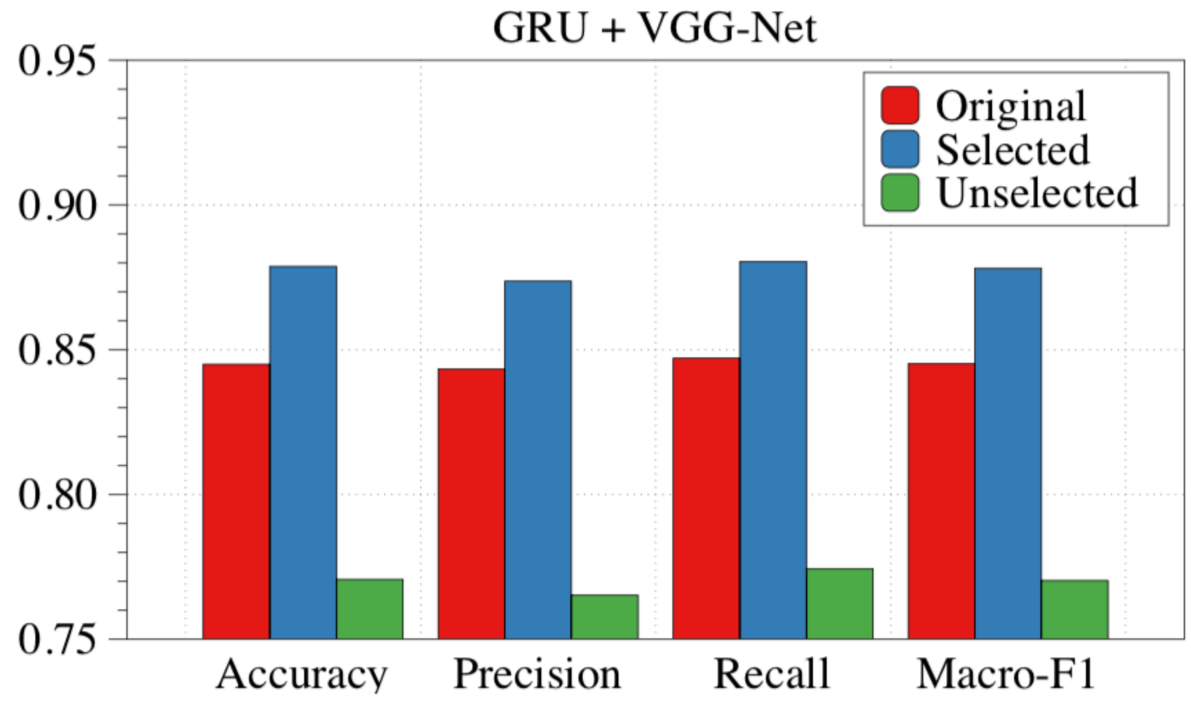


Figure 3: Comparison of models trained on original posts, selected posts, and unselected posts.

Experiments

Analysis

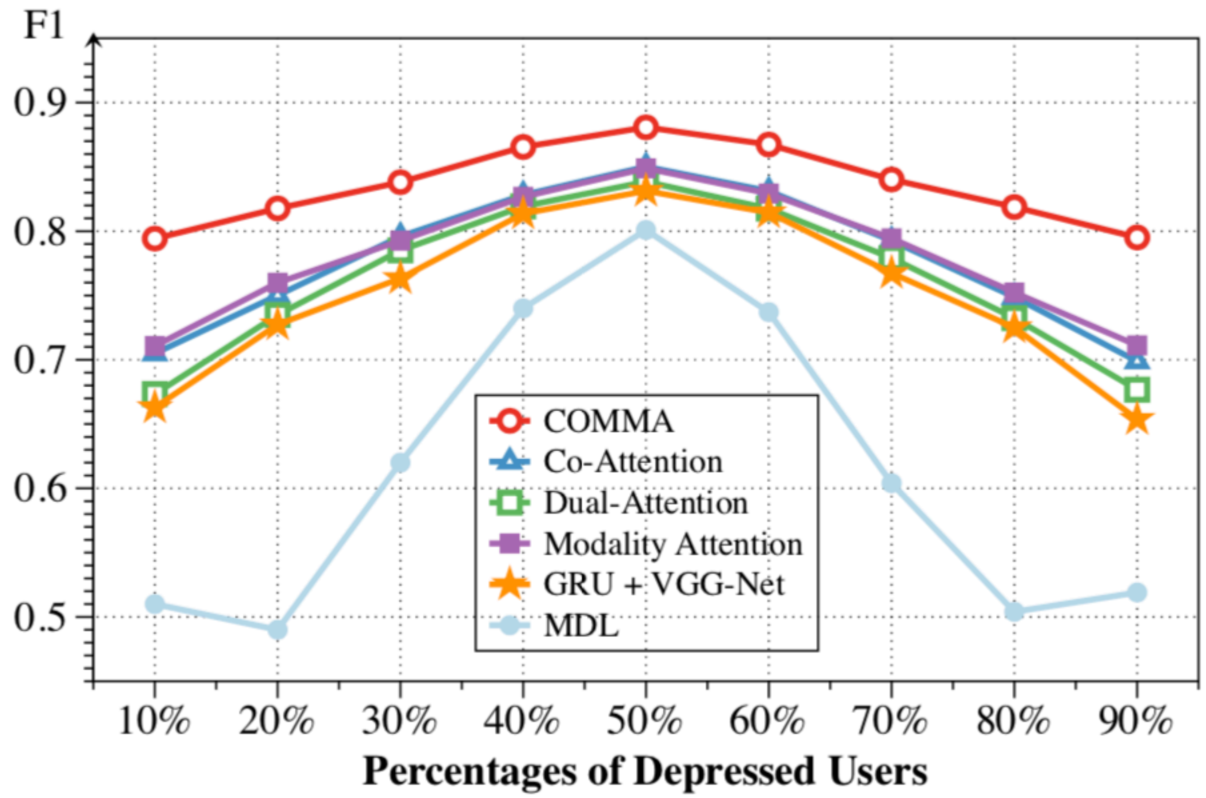


Figure 4: Comparison of the models trained on the datasets with different percentages of depressed users. The total number of users is 1,500.

Experiments

Analysis

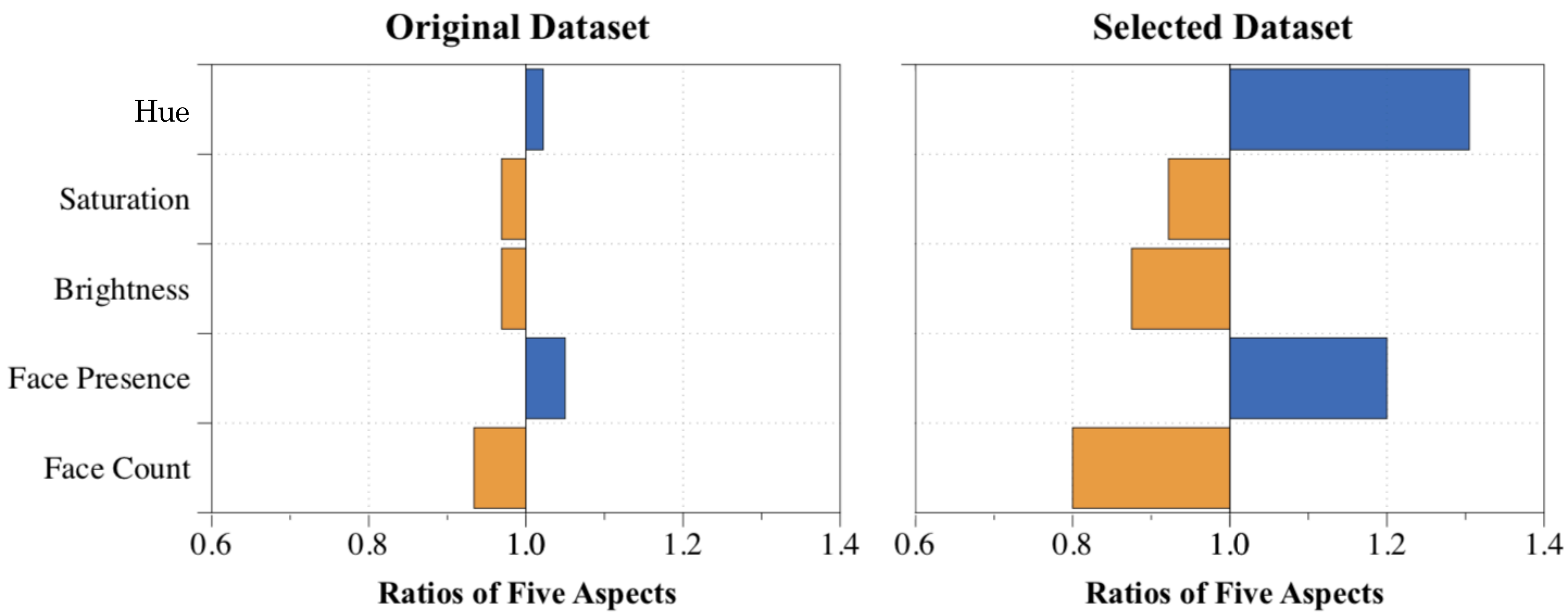


Figure 5: Comparison of original and selected posts. The y-axis values show the five aspects of each image, and the x-axis values are the ratios of these five aspect values of depressed users to those of non-depressed users.

- Consistent with those reported in (Reece and Danforth 2017)

Experiments

Analysis

Dataset	Top words (by frequency)
Selected data of depressed users	bad, cancer, insurance, hate, medical, pain, cost, mental, ...
Unselected data of depressed users	people, online, time, know, life, free, school, weight, work, ...
Original data of non-depressed users	wow, idk, like, party, gotta, funny, ◡̈, honestly, team, :) ...

Table 3: Example words arranged in descending order of word frequency.

Outline

1. Introduction and Motivation
2. Approach
3. Experiments
4. Conclusion

Conclusion

Conclusion

- We study the problem of detecting depression by incorporating textual and visual information.
- We propose a novel multi-agent reinforcement learning method, COMMA, to achieve this task, in which text and image selectors cooperatively extract indicator content.
- Experimental results for the depression benchmark show that COMMA can significantly improve performance.

Thank You