MG-PRIFAIR: MULTIMODAL REVIEW GENERATION

with Privacy and Fairness Awareness

We introduce MG-PriFair, a multimodal neural-based framework, which generates personalized reviews with privacy and fairness awareness. We propose a novel differentially private (dp)-embedding model for training privacy guaranteed embeddings and an evaluation approach for sentiment fairness in the food-review domain. Experiments show that MG-PriFair is capable of generating plausibly long reviews while controlling the amount of exploited user data and using the least sentiment-biased word embeddings. To the best of our knowledge, we are the first to bring user privacy and sentiment fairness into the review generation task. The dataset and source codes are available at https://github.com/ReML-AI/MG-PriFair.

1

Introduction to the model:

- **Privacy Controller:** controls the amount of user information.
- **Fairness Controller:** evaluates the sentiment bias (fairness) of word embeddings to be used in the generation model.
- **PRGen Model**: a Personalized Review Generation Model receives receives as input an image, user, and entity e, and outputs a review document.

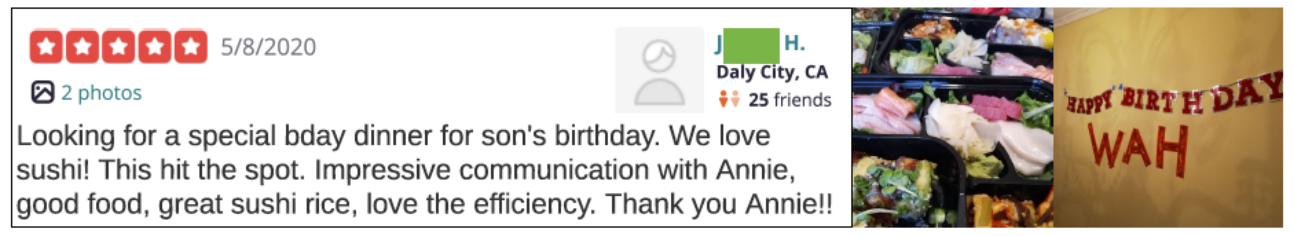

Figure 1: An example of online review from Yelp.com that contains personal information.

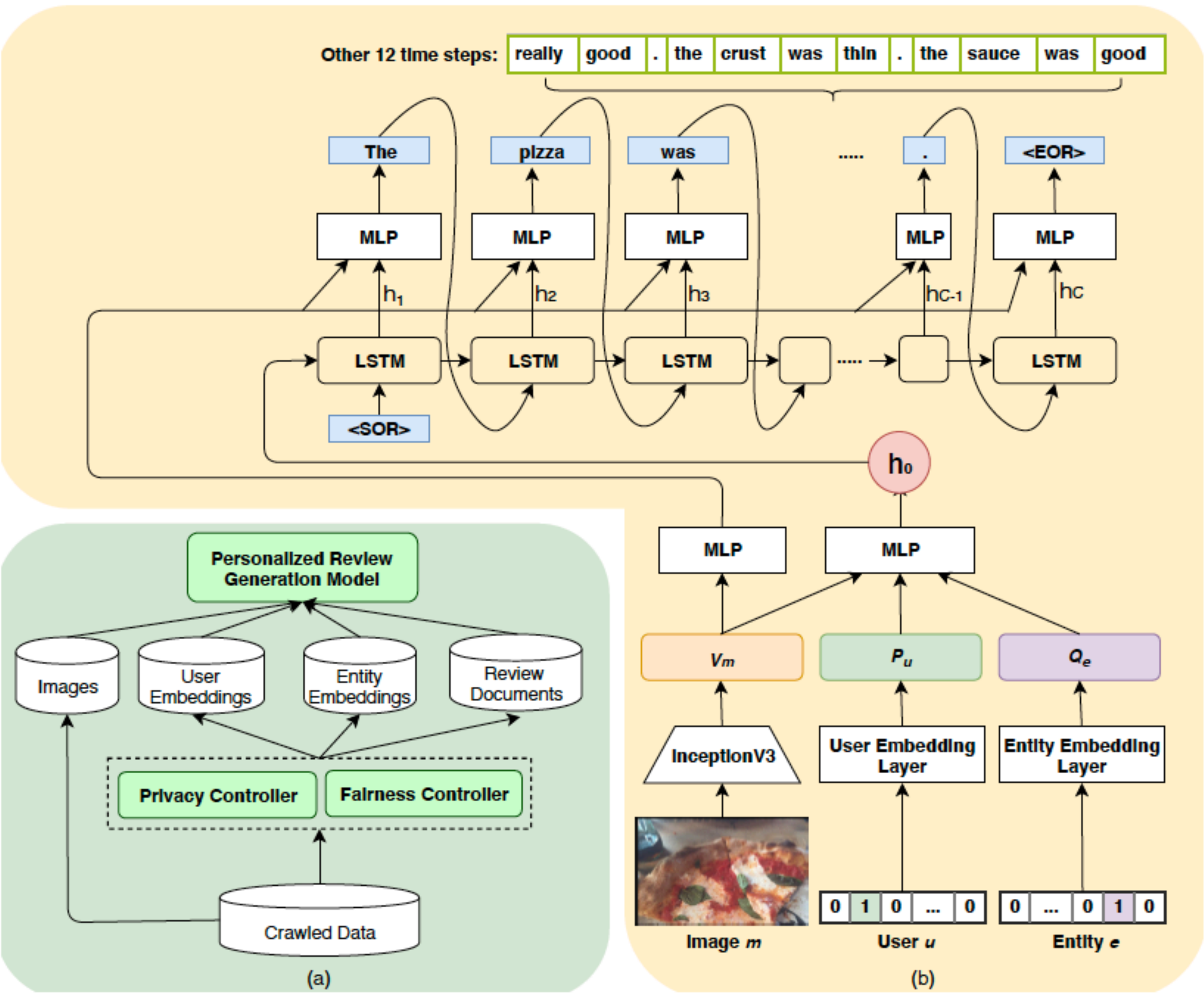

Figure 2: a) Architecture design of MG-PriFair which includes privacy and fairness controllers and a generation model. b) Our proposed personalized review generation model (PRGen).

2

Dataset: We construct a new dataset from Yelp.com that contains restaurant reviews for seven different English speaking cities.

(a) for Each City.

City	#reviews	#users	#entities	#images
Charlotte	16,356	6,732	880	24,523
Edinburgh	1,763	987	543	2,586
Las Vegas	80,152	39,415	868	126,389
London	7,078	3,878	946	10,651
Phoenix	31,830	15,681	947	46,728
Pittsburgh	14,016	5,226	892	21,420
Singapore	3,151	1,091	891	4,958
Total	154,346	69,472	5,967	237,255

(b) for Train/Validation (Val)/Test Sets.

	#tuples	#images	#users	#entities
Train	1,055,591	218,347	66,782	5,933
Val	45,844	9,466	7,328	2,949
Test	45,801	9,442	7,346	2,900

Table 1: Statistical information of the Dataset

Image	Ground-truth	*MG-PriFair*
	and while you 're at it , and speaking of pho again ... lobster pho . this dish is an absolute memorable dish .	i had the lobster pho and the lobster pho . both were really good . the lobster pho was delicious .
	great ramen place in vegas chinatown . i have not been to too many ramen places but this was a nice small place . they had a page of different options and portions .	the broth was delicious and the noodles were cooked perfectly . the broth was flavorful and the noodles were cooked perfectly . the broth was rich and flavorful .

Figure 3: Examples of given images and generated reviews in comparison to ground-truth of MG-PriFair.

3

Fairness Evaluation:

A new R-WEAT list is proposed to evaluate fairness of Word Embedding in Food Review domain.

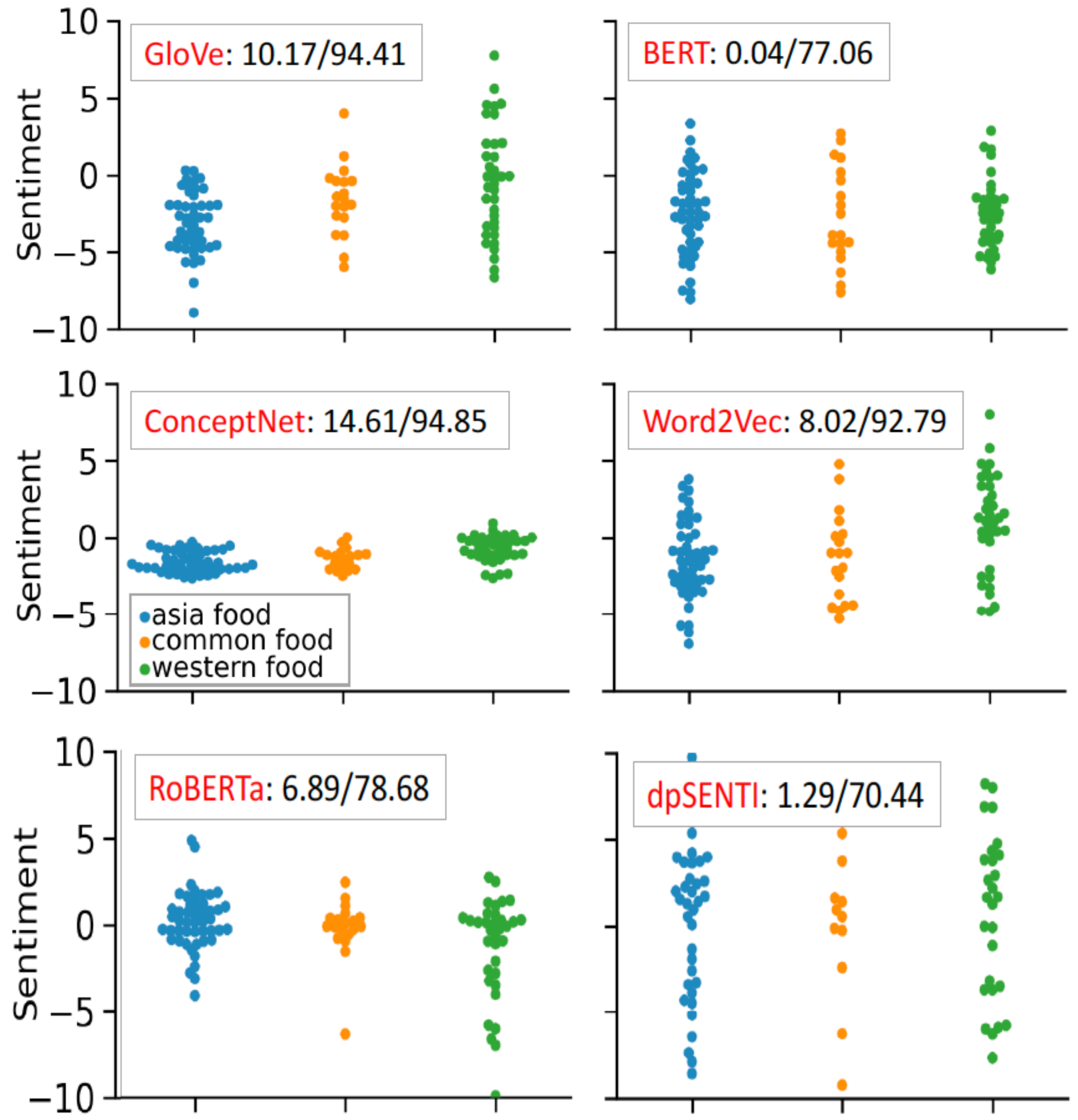

Figure 4: Fairness evaluation based on R-WEAT list for different word embeddings models. F-bias value (F) and classification accuracy (A) for each word embeddings model are reported in the form of F/A. Here, we proposed **dpSENTI** , a differential private embedding optimized for sentiment classification task. Visualization method was inspired by Speer (2017b).

4

Evaluations:

- Generation evaluations
- Qualitative evaluations

Generation evaluations: Comparison between our model (PRGenG-PY) and the baselines in image caption (ShowNTell) and review generation (MRG) in terms of Bleu-4, sentiment polarity, subjectivity, grammar errors (GramErr) and sentences' length (AvgLen). The superscript * marks the metrics in which the lower value the better. POS, NEG, Avg and GT stand for positive, negative, average and groundtruth, respectively.

Model	Bleu-4	*Polarity* POS	NEG*	#Zeros*
ShowNTell	2.58	0.44	-0.22	1302
MRG	1.07	0.41	**-0.29**	910
PRGen$_G$-PY	**2.77**	**0.55**	-0.21	**424**

Model	*Subjectivity* Avg	#Zeros*	GramErr*	AvgLen (GT:15)
ShowNTell	0.51	1157	3.86	31
MRG	**0.65**	760	**0.46**	50
PRGen$_G$-PY	**0.65**	**362**	0.94	10

Table 2: Comparison with image captioning baselines.

Qualitative evaluations: Here we have 100 images and reviews, in which half of reviews generated by our personalized review generation model. Each participant votes all images, to justify if a review was written by human or machine.

		Predicted Result Human	Machine
Actual Result	Human	62%	38%
	Machine	49%	51%

Table 3: Voting results of User Study.

Agency for Science, Technology and Research SINGAPORE

UMEÅ UNIVERSITY

Authors: Xuan-Son Vu (UMU), Thanh-Son Nguyen (A*STAR), Duc-Trong Le (VNU-UET), Lili Jiang (UMU)
Emails: {sonvx, lili.jiang}@cs.umu.se; nguyents@ihpc.a-star.edu.sg, trongld@vnu.edu.vn

VNU University of Engineering and Technology