Paper Number: 3066

ICASSP2021 TORONTO Canada June 6-11, 2021 Metro Toronto Convention Centre

COOPNET: MULTI-MODAL COOPERATIVE GENDER PREDICTION IN SOCIAL MEDIA USER PROFILING

User Profiling

Presenter: Kaixi Hu

Date: June, 2021

* Corresponding author. Email: cathylilin@whut.edu.cn

Source code: https://github.com/WUT-IDEA/COOPNet

Lin Li[1,*] Kaixi Hu[1] Yunpei Zheng[1] Jianquan Liu[2] Kong Aik Lee[3]

[1] School of Computer Science and Technology, Wuhan University of Technology, China
[2] Biometrics Research Laboratories, NEC Corporation, Japan
[3] Institute for Infocomm Research, A*STAR, Singapore

武漢理工大學 Wuhan University of Technology

Orchestrating a brighter world NEC

Institute for Infocomm Research I²R

01 INTRODUCTION

➢ Background

Information asymmetry in user generated contents：

- Semantic complementarity.

For user A:

win a match ticket.

A football match.

By taking both them into consideration

User A watched a football match on scene.

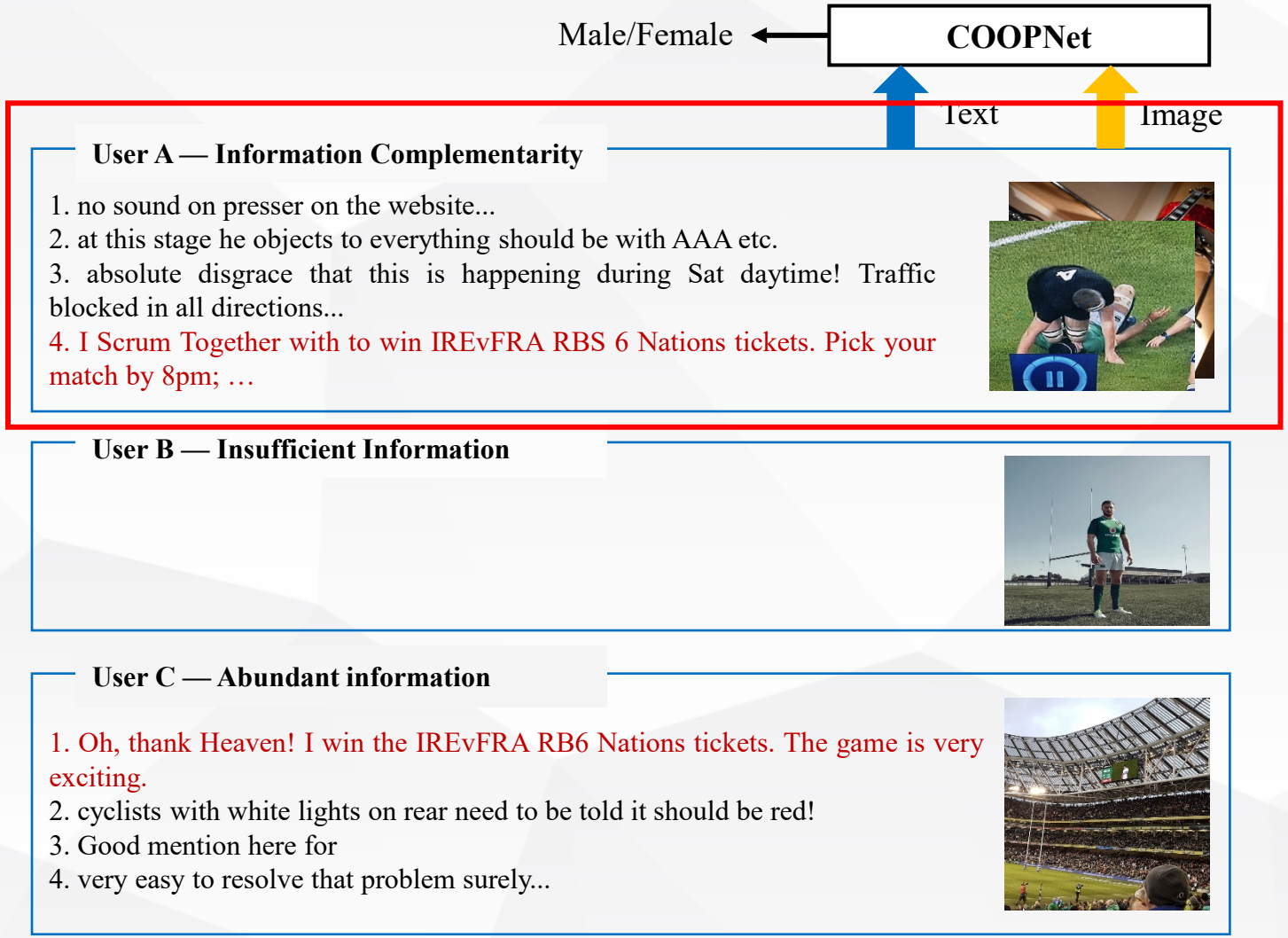


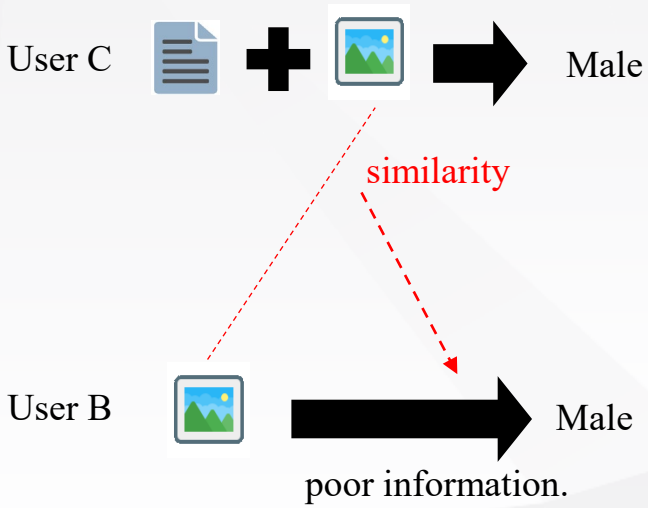
Figure: Pairs of text-image contents from users A, B and C.

01 INTRODUCTION

➢ Background

Information asymmetry in user generated contents：

- Insufficient information in unilateral modality



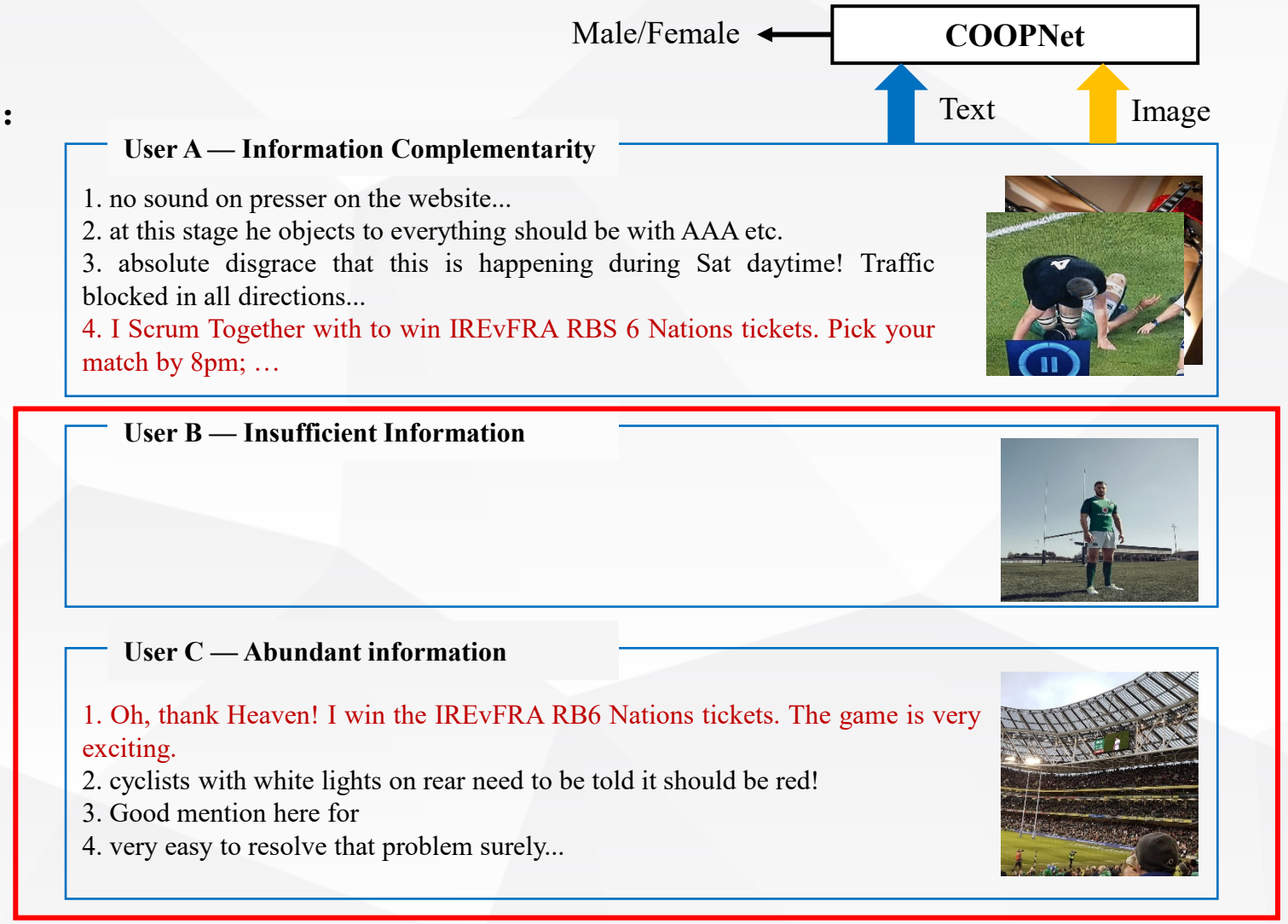


Figure: Pairs of text-image contents from users A, B and C.

01 INTRODUCTION

➢ Related Works

Single-modal Prediction

- **1. Various types of data:** a variety of textual data [1] new annotated corpus [2]
- **2. Plentiful features:** lexical features [3] behavioral features [4]
- **3. model architectures:** Bi-GRU and CNN [5] a ensemble LSTM model [6]

Insufficient information

Multi-modal Prediction

- **1. Fusion based on feature engineering:**
  - Weighting: [7], [8], [9]
  - Concatenation: [10]
  - Rule : [11]
- **2. Fusion based on automatic feature learning:** Weighting: [12], [13]

Simple fusion operator: concatenation or weighting

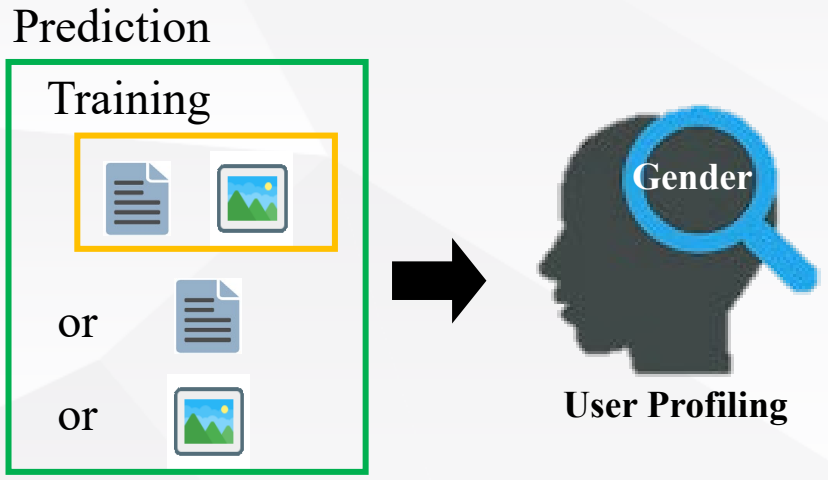
[1] John D. Burger, John C. Henderson, George Kim, and Guido Zarrella. 2011. Discriminating Gender on Twitter. In EMNLP, 1301–1309.
[2] Daniel Preotiuc-Pietro, Vasileios Lampos, and Nikolaos Aletras. 2015. An analysis of the user occupational class through Twitter content. In ACL, 1754–1764.
[3] Wen Li and Markus Dickinson. 2017. Gender Prediction for Chinese Social Media Data. In RANLP, 438–445.
[4] Zhe Zhao, Zhiyuan Cheng, Lichan Hong, and Ed Huai-hsin Chi. 2015. Improving User Topic Interest Profiles by Behavior Factorization. In WWW. ACM, 1406–1416.
[5] Jaeyong Kang, Hongseok Choi, and Hyunju Lee. 2019. Deep recurrent convolutional networks for inferring user interests from social media. J. Intell. Inf. Syst. 52, 1 (2019), 191–209.
[6] Dong Zhang, Shoushan Li, Hongling Wang, and Guodong Zhou. 2016. User Classification with Multiple Textual Perspectives. In COLING. ACL, 2112–2121.
[7] Giovanni Ciccone, Arthur Sultan, Lea Laporte, Elod Egyed-Zsigmond, Alaa Alhamzeh, and Michael Granitzer, Stacked gender prediction from tweet texts and images. in CLEF, 2018, vol. 2125.
[8] Moniek Nieuwenhuis and Jeroen Wilkens, Twitter text and image gender classification with a logistic regression n-gram model. in CLEF, 2018, vol. 2125.
[9] Eric Sadit Tellez, Sabino Miranda-Jimenez, Daniela Moctezuma, et al., Gender identification through multi-modal tweet analysis using microtc and bag of visual words. in CLEF, 2018, vol. 2125.
[10] Sebasti´ an Sierra and Fabio A. Gonzalez, “Combining textual and visual representations for multimodal author profiling,” in CLEF, 2018, vol. 2125.
[11] Matej Martinc, Blaz Skrlj, and Senja Pollak, Multilingual gender classification with multi-view deep learning. in CLEF, 2018, vol. 2125.
[12] Nils Schaetti, Character-based convolutional neural network and resnet18 for twitter author profiling. in CLEF, 2018, vol. 2125.
[13] Takumi Takahashi, Takuji Tahara, Koki Nagatani, et al., Text and image synergy with feature cross technique for gender identification. in CLEF, 2018, vol. 2125.

02 OUR FRAMEWORK

➢Task Description

Input: texts, images from a user and a extra sentiment dictionary

Output: the user's gender



➢ Model Architecture

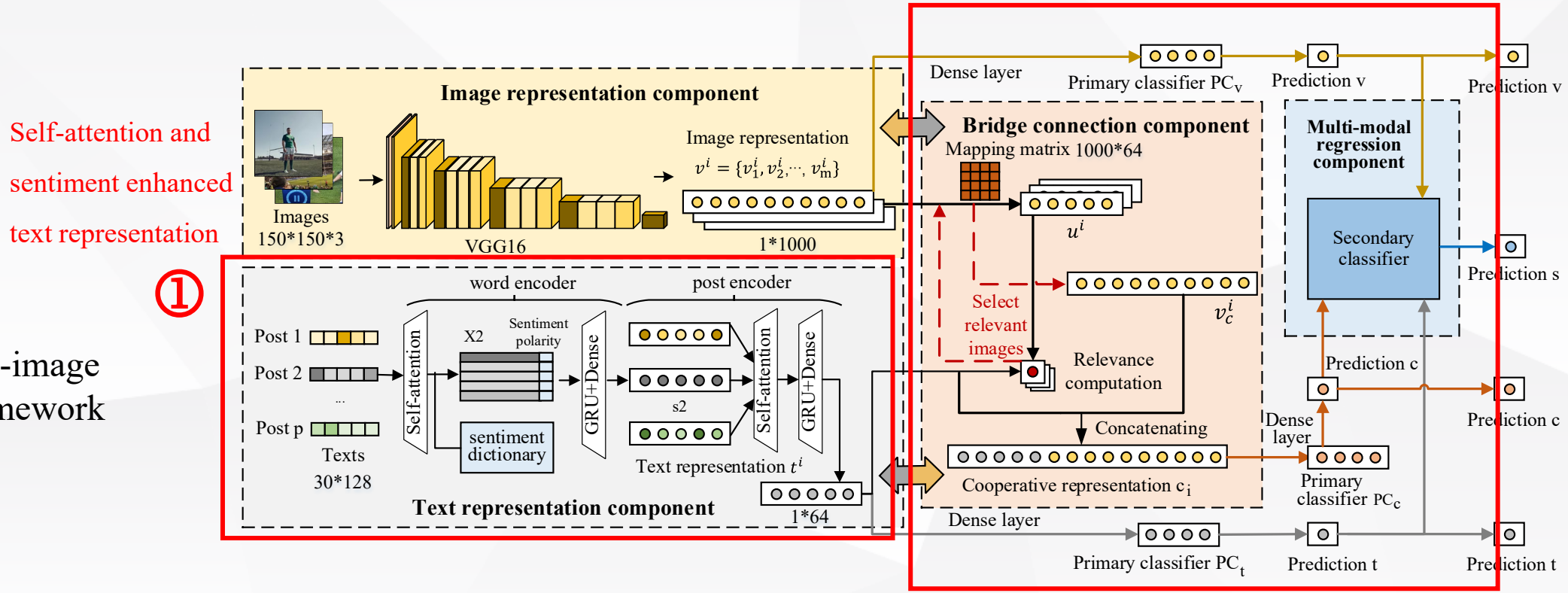


Figure: The text-image cooperation framework COOPNet.

02 OUR FRAMEWORK

➢ Enhanced Text Representation

- Self-attention can figure out relationship w.r.t. distance [2].

- Texts are easily to be enhanced by introducing external knowledge, such as sentiment [1].

Each word embedding is concatenated with its sentiment polarity:

$$X_2 = Concat[X, P]$$

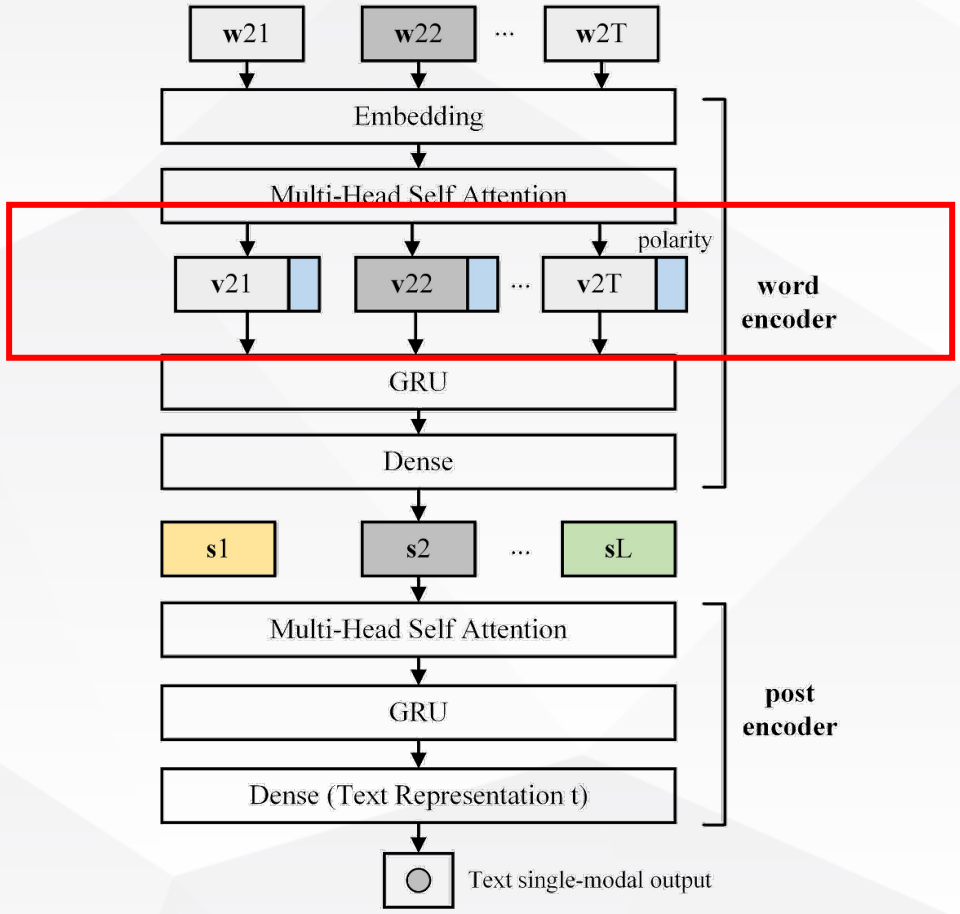


Figure: Sentiment Enhanced Text Representation.

[1] Yunpei Zheng, Lin Li, Jianwei Zhang, Qing Xie, and Luo Zhong. 2019. Using Sentiment Representation Learning to Enhance Gender Classification for User Profiling. In Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data. Springer, 3–11.
[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In NIPS. 5998–6008.

02 OUR FRAMEWORK

➢ Semantic Complementarity

- Bridge Connection

a mapping matrix W_{m} → A cooperative space.

$$\alpha_j^i = \frac{\exp(t^{i^T} u_j^i)}{\sum_{j=1}^{m} \exp(t^{i^T} u_j^i)}$$

Texts are employed as a north star to align relevant images $v_c^i = \sum_{j=1}^{m} \alpha_j^i v_j^i$

→ the cooperative representation $c^i = Concat[t^i, v_c^i]$

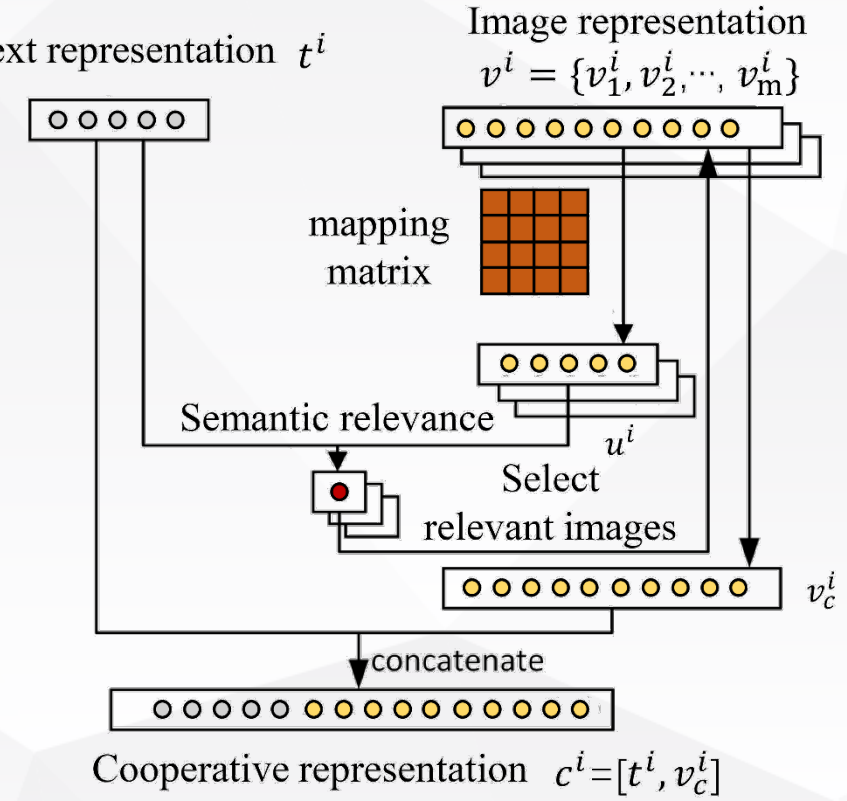


Figure: The bridge connection component.

- Multi-modal Regression

Further enhance the ability of information.

$$\hat{y}^i = \frac{1}{1 + exp(-w_t^T \hat{y}_t^i - w_v^T \hat{y}_v^i - w_c^T \hat{y}_c^i + b_c)}$$

03 EXPERIMENTS

➢ Setting

- Dataset

  3000 and 1900 samples with 1:1 male-to-female ratio are divided for training and testing, which follows the PAN-2018 competition.

  Metrics: AUC, F1-Measure

Dataset: https://pan.webis.de/clef18/pan18-web/author-profiling.html

- Preprocess

  1. Zero matrix fill up the missing image and resize all images into the size of (150,150,3).
  2. The maximum length of a post is limited to 50 and the word embedding dimension to 300.
  3. Average the RGB value of all images to normalize the images further

- Parameters

  The learning rate, attention head, text and image representation dimension are set as 0.01, 2, 64 and 1000, respectively.

Source code: https://github.com/WUT-IDEA/COOPNet

03 EXPERIMENTS

➢Baselines

Table: Baselines divided by fusion and feature engineering

Model	Text	Image	Fusion(Early/Late)	Feature Engineering	
Ciccone et al. [1, 2]	N-grams, TF-IDF	6 classifiers	LinearSVC(Late)	Yes	State-of-the-art
Nieuwenhuis et al. [1]	N-grams, Glolve	13 features	LR(Early)	Yes	
Tellz et al. [1]	N-grams	Image caption	Weighted average(Late)	Yes	
Aragon et al. [1]	N-grams	VGG16	SVM(Early)	Yes	
Sierra et al. [1]	Bag-of-Words	ResNet50, VGG16	Concat(Early)	Yes	
Kerner et al. [1]	N-grams, Stylistic features	SIFT, Color, VGG	Weighted average(Late)	Yes	
Martinc et al. [1]	CNN, TF-IDF	Face detection	three conditions(Late)	Yes	
Stout et al. [1]	TF-IDF, N-gram, RNN	CNN, Pool	Weighted average(Late)	Yes	
Patra et al. [1]	Word2vec, TF-IDF, LSA, LDA	Image caption	SVM(Early)	Yes	
*Takahashi et al. [1, 3]	RNN	VGG16	Direct-product(Early)	No	State-of-the-art
Schaetti et al. [1]	Character based CNN	ResNet18	Average probabilities(Late)	No	
COOPNet(ours)	Self-attention, GRU	VGG16	Bridge(Early), LR(Late)	No	

- Most of existing methods are proposed in PAN-2018 competition. (https://pan.webis.de/clef18/pan18-web/author-profiling.html)
- Pardo et al. summarize some top methods and results in the PAN-2018 competition [1].
- For fair comparison, we reproduce Takahashi et al.'s method and strip out extra streaming tweets used for pre-trained word embeddings.

[1] Francisco M. Rangel Pardo, Paolo Rosso, Manuel Montes-y-G´ omez, Martin Potthast, and Benno Stein, "Overview of the 6th author profiling task at PAN 2018: Multimodal gender identification in twitter," in CLEF, 2018, vol. 2125.
[2] Giovanni Ciccone, Arthur Sultan, Lea Laporte, Elod Egyed-Zsigmond, Alaa Alhamzeh, and Michael Granitzer, "Stacked gender prediction from tweet texts and images," in CLEF, 2018, vol. 2125.
[3] Takumi Takahashi, Takuji Tahara, Koki Nagatani, Yasuhide Miura, Tomoki Taniguchi, and Tomoko Ohkuma, "Text and image synergy with feature cross technique for gender identification," in CLEF, 2018, vol. 2125.

03 EXPERIMENTS

➢ Results

- COOPNet outperforms most baselines and achieves the best automatic feature learning based methods.
- COOPNet shows a outstanding result when only image is fed into model.

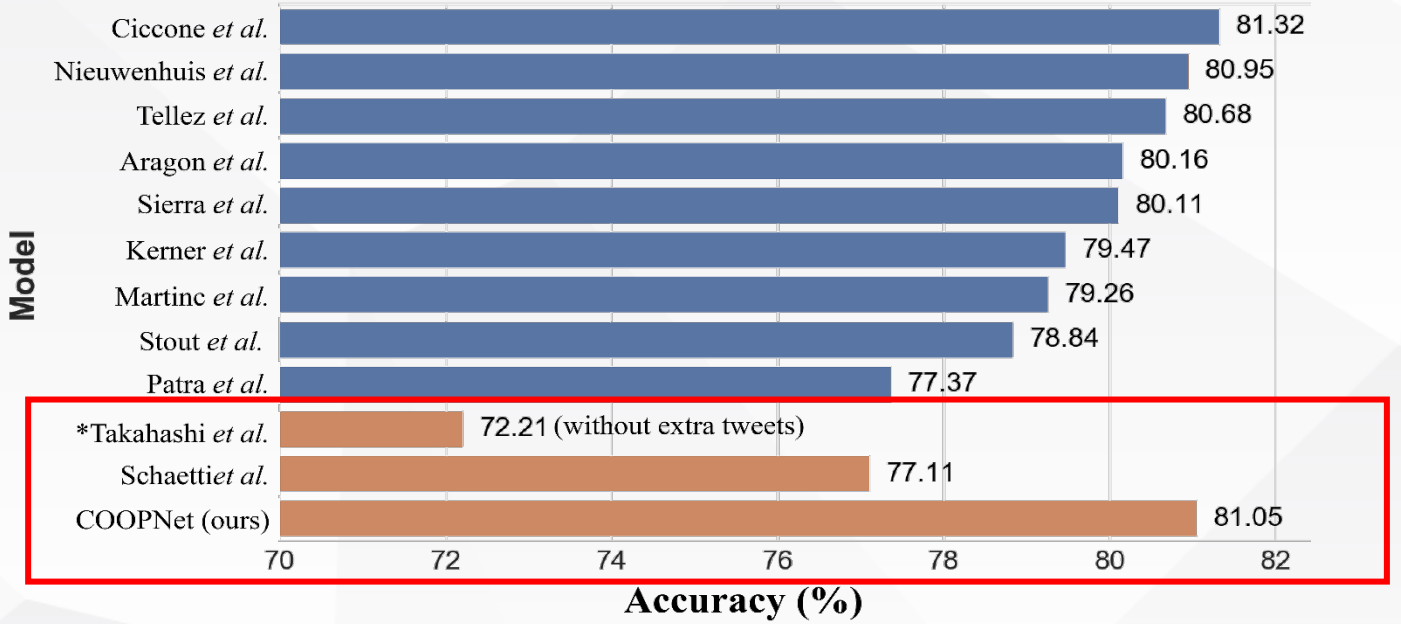


Figure: Accuracy comparison in multi-modal prediction. The blue bars present feature engineering based methods and the orange bars present the automatic feature learning based methods .

Table: Comprehensive accuracy comparison between COOPNet and Ciccone et al's method in prediction。

Model	Image Input	Text Input	Multi-modal input
COOPNet(ours)	**80.63**	80.73	81.05
Ciccone et al.	69.63	80.74	81.32

03 EXPERIMENTS

➢ Ablation Analysis

- Through the bridge connection component, a lot of text knowledge are transferred to the image component and relatively less image knowledge are transferred to the text component.

- The multi-modal regression component can learn some extra information from the soft probabilities since each probability is generated from different initial conditions

Table: Evaluation on various components.

① ②

Trainning	Component	Accuracy(%)	AUC	F1
single-modal	text (PC_t)	79.31	0.8301	0.8047
	text-sentiment (PC_t)	77.57	0.8112	0.7649
	image (PC_v)	75.57	0.7757	0.7292
multi-modal	text (PC_t)	**80.73**	0.8696	**0.8142**
	image (PC_v)	80.63	0.8486	0.8108
	bridge (PC_c)	80.57	**0.8718**	**0.8142**
	$COOPNet(PC_s)$	**81.05**	**0.8730**	**0.8150**

04 CONCLUSION AND FUTURE WORK

➢ Conclusion

- The problem of information asymmetry generally exists in user generated contents.
  (semantic complementarity between different modalities, insufficient information in unilateral modality)
- Enrich semantic: sentiment and self-attention enhanced text representation.
- Multi-modal cooperation: bridge connection and multi-modal regression.

➢ Future Work

THANKS

If you are interested in our work, more details can be found in our paper:

Lin Li*, Kaixi Hu, Yunpei Zheng, Jianquan Liu, Kong Aik Lee. COOPNet: Multi-modal Cooperative Gender Prediction in Social Media User Profiling. 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'21), 6 June, 2021, Toronto, Ontario, Canada