Fusion of Multi-modal Information of User Profile across Social Networks for User Identification

Cuicui Ye
College of Computer Science and Technology
Harbin Engineering University
Harbin, China
yecuicui@hrbeu.edu.cn

Jing Yang
College of Computer Science and Technology
Harbin Engineering University
Harbin, China
yangjing@hrbeu.edu.cn

Yan Mao
College of Information Engineering
Harbin University
Harbin, China
maoyan@hrbeu.edu.cn

***Abstract*—User identification across social networks uses a variety of user information to determine whether two accounts from different social networks belong to the same user. The most intuitive method is to use user profiles to solve user identification across social networks. How to effectively learn the characteristics of user profiles is crucial. In this paper, we propose a model for user identification across social networks based on multi-modal information fusion of user profiles. First, use the pre-trained model in deep learning to obtain the text feature vector and image feature vector of the user profile. We fuse the multi-modal feature vectors in the user profile. Then, the feature vector sequence is fed into the bidirectional LSTM model with an attention mechanism. Finally, the probability result of user identification is obtained through the fully connected layer and softmax layer. Experimental results show that the suggested technique outperforms state-of-the-art baselines when evaluated on three real-world datasets. Our approach outperforms solutions based on image pixel-level comparison when it comes to user identification challenges, thanks to the semantic feature mining of photos in user profiles.**

Keywords—Across social networks, User identification, Pre-trained model, user profile

I. INTRODUCTION

With the rapid development of the Internet, social networks have become an indispensable part of people's daily lives. People enjoy various services on multiple social networking platforms for different purposes. The goal of user identification is to accurately identify the same person in different social networks [1]. User identification across social networks, also known as User Identity Linking (UIL), has attracted increasing attention due to its important value in recommender systems [2] and network security [3].

Existing works measure the similarity of users across social networks through user profiles[4] [5], network structures[6] [7], user-generated content[8] [9], or action trajectories[10] [11], and then use the similarity to determine whether two users across social networks are the same person. These tasks can identify some of the same users. However, social networks are independent of each other, and social network data is noisy. Accessible properties are inconsistent and incomplete across different social networks. For example, some social networks restrict user content and network access (Weibo users can set the visible range of published content, while WeChat does not allow the display of user-friend relationships.).

Similarities in profiles can be used for user identification. However, existing research faces two challenges:

(1) User attributes are missing or false. When users register accounts in multiple social networks for privacy protection, the user attributes may be empty or forged except for the user name and user avatar. However, it might be challenging to take advantage of these features because different profiles may contain inaccurate, incomplete, or false information.

(2) Insufficient mining of semantic features. For example, existing methods apply string-matching metrics such as Jaro-Winkler distance and Levenshtein distance to match user names across social networks. Not taken into account are the semantic characteristics of text attributes like usernames.

We propose an algorithm for the Fusion of Multi-modal Information of User Profile across Social Networks for User Identification to address these two issues.

Firstly, users can configure their username and user avatar, which are basic profile features that are publicly viewable, among other profile information. Because 59% of users prefer to use the same username on numerous social media platforms [12], the username is a crucial component of user identification. An essential visual component of social media users' profiles is their avatars. Shu et al. [13] emphasized the relationship between the user's personality, actions, and behavior and the selection of profile images. Ranaldi et al. [14] demonstrated that using images in profiles can match the profiles of users in different SNs with high performance. Therefore, we select user names and user avatars in user profiles for cross-social network user identification to overcome the problems of missing user attributes and falsehoods.

Secondly, with the development of deep learning, the extraction of deep semantic features becomes possible. Huang et al. [15] extract document-level semantic feature vectors from UGC based on the BERT model. Du et al. [16] used deep learning models to extract semantic features from user-generated text to select similar user pairs based on similarity in writing styles. Ye et al. [17] apply the multi-lingual BERT pre-

This work was supported in part by the National Natural Science Foundation of China under Grant no.61672179..

training model to extract the deep semantic features of user names. Inspired by the above research, this paper applies the BERT pre-training model to extract the semantic features of user names.

In addition, user profile information containing text and images is multi-modal information, so the problem of multi-modal information fusion becomes the key to solving the problem of cross-social network user identification based on user profiles. In recent years, there has been rapid development of large-scale pre-training frameworks that can extract multi-modal representations in a unified form and achieve good performance when transferred to downstream tasks [18]. To solve the insufficient semantic feature mining of user profile information and the representation and fusion of multi-modal information, we apply the clip model [19]. The Clip model can not only extract the semantic features of usernames and user images but also make the fusion of different modal feature vectors possible. Overall, the main approaches are as follows:

1) We use the clip model to extract semantic features from the user avatars, and then perform multi-modal feature fusion to obtain semantic features with the user's avatar.

2) We constructed a user identification model with an attention machine based on bidirectional LSTM and used the multi-modal fusion classification features of user profiles to identify users across social networks.

3) We conduct experiments on three real-world datasets and the results demonstrate the superior performance of our method on user identification tasks.

II. Relate Works

A. User identification based on user profile

User profile attributes include username, avatar, biography, gender, country, postcode, etc. Some scholars have conducted research using only usernames. Using the longest common substring, Li et al. [20] created a UISNUD model that assessed the similarity between a username and a display name. By making use of username redundancy on social networking sites, Zafarani et al. [21] developed a Mobius technique based on naming patterns that functioned effectively across several networks. Li et al. [22] used machine learning approaches to identify users based on hand-crafted features after analyzing users' distinctive naming practices and using information redundancy of usernames to generate features. These models do not perform well with Chinese datasets, but they are good at detecting anchor users with English usernames. Li et al. [23,24] employ the display name embedding method of Chinese character morphology and phonetic information to address the problem of user identification in Chinese social networks. These techniques are limited by in-depth topic expertise and depend on manually constructed feature extraction of usernames.

Some scholars study the problem of user identification across social networks based on multi-dimensional attribute information. Wang et al. [25][26] considered the multi-dimensional attributes of the profile, but did not consider the user's avatar. Wang et al. [26] choose Levenshtein distance to measure the similarity of usernames, and combine the text information in the profile into a bag-of-words model. Then the cosine distance is used to measure the similarity of the two profiles. Shu et al. [13] applies word2vec to vectorize user attributes (username, city of residence, interest preferences) to extract word-level semantics. Some scholars use Levenshtein distance or Jaro-Winkler Distance to calculate the username similarity[27][28][29]. To capture the similarity between profile avatars, they use OpenFace to crop the image to extract the face, and calculate the distance between the two image vectors distance. But portraits are not the only choice for user avatars. This method is not suitable for all situations.

B. Development of pre-trained models

The combination of pre-trained models and fine-tuning effectively solves the problem of training time and economic costs. Model-level pre-training models such as ELMO[30], GPT[31], BERT[32], and CLIP[19] have been derived. First, the pre-training model is pre-trained on a large-scale general data set using a self-supervised approach, so that the model can learn a general representation before moving to downstream tasks, and then fine-tune it on a proprietary data set to obtain task-specific knowledge. BERT is a pre-trained language model that achieves state-of-the-art results in the text domain. CLIP uses natural language as supervision to improve the image classification effect and uses contrastive learning methods to promote the matching ability of images and text. Inspired by the above work, we applied Bert and clip to extract the semantic features of text and images in user profile information and obtained the multi-modal feature vector of the user profile.

III. The Proposed Model

To solve the problem of user identification based on multi-modal attributes of user profiles, we propose a multi-modal feature fusion method combined with pre-trained models. Fig.1 depicts the overall framework of the model, detailing each component that makes up the architecture.

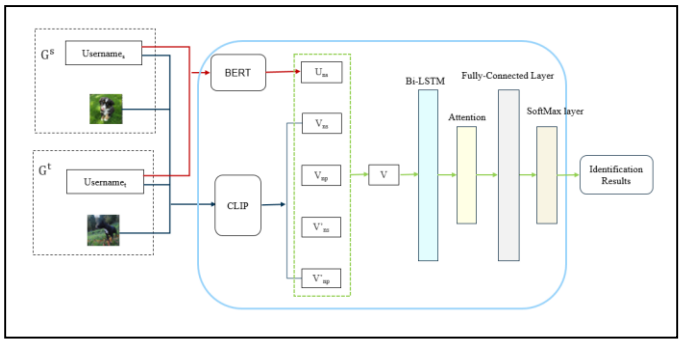

Fig. 1. Overall frame structure.

Feature extraction and feature fusion make up the two primary sections of our model. In feature extraction, we use the Cross-Encoder structure of pre-trained Bert to concatenate the two user names and send them to the cross-encoder to obtain a CLS classification feature. Then use the clip model to extract the semantic features of user text and images. Use the Text Encoder in the clip to extract the semantic feature vector of the text, and use the Image Encoder in the clip to extract the semantic feature vector of the image. Finally, five feature vectors were obtained: the text semantic feature vector and avatar semantic feature vector from the two platform users extracted by the clip model,

as well as the classification label feature vector output by Cross-Encoder. We splice these 5 feature vectors into a sequence and send it to the bidirectional LSTM with an attention mechanism for feature fusion. The feature fusion part includes two layers of feature fusion, namely the fusion of semantic feature vectors and the fusion of classification features. After fusing semantic features, we add an attention mechanism to obtain a classification feature representation. Finally, the classification feature vector is sent to the fully connected layer, and then the probability output is obtained through the softmax layer.

A. Semantic feature extraction based on Corss-Encoder model

The pre-trained language model abandons the previous modeling method of character and word granularity and directly models sentences. And capture the long-range dependence of sentences through the attention mechanism. This solves the prior challenges in deep neural network sentence modeling. Multiple text pairings can be encoded and judged simultaneously by the cross-encoder. The model receives several texts as an input as a whole, learns to represent the relationships between various text pairings, and produces correlation scores or labels between the texts. In our model, the Corss-Encoder passes both usernames to the Transformer network simultaneously. We get a <CLS> classification label vector V_B^{name} with the semantic features of the username.

B. Semantic feature extraction based on CLIP

The full English name of CLIP is Contrastive Language-Image Pre-training, which is a pre-training method or model based on contrasting text-image pairs. CLIP is a multi-modal model based on contrastive learning. The training data is a text-image pair: an image and its corresponding text description. Through contrastive learning, the model can learn the matching relationship of the text-image pair. CLIP includes two models: Text Encoder and Image Encoder. Text Encoder is used to extract text features and uses the text transformer model in NLP; while Image Encoder is used to extract image features and uses the CNN model or vision transformer. Since the CLIP model can obtain the semantic feature vectors of text and images, we use Text Encoder to obtain the semantic feature vector V_C^{ns} *and* $V_C^{nt}$of the username and Image Encoder to obtain the semantic feature vector V_C^{ps} *and* V_C^{pt} of the avatar.

C. Multimodal Fusion

Multimodal representation learning vectorizes the semantic information contained in multimodal data through embedding. Then, the models and features between different modalities are integrated through multi-modal fusion. Feature vectors from different modalities can be integrated through simple operations such as concatenation and weighted summation. Our task is to identify whether two users are the same person, so we extract differential features. Usernames were compared with differentiating features using cross-encoder and clip, and images were compared with differentiating features with clip. We use difference generation to fuse the four feature vectors generated by the username and avatar. Eq. (1)-(3) is the formula expression of the fusion vector.

$$V_{ps-pt} = v_C^{ps} - v_C^{pt} \tag{1}$$

$$V_{ns-nt} = v_C^{ns} - v_C^{nt} \tag{2}$$

$$V_{fusion} = \left(v_C^{ns} + v_C^{ps}\right) - \left(v_C^{nt} + v_C^{pt}\right) \tag{3}$$

$$V = \left(V_B^{name}, V_C^{ns}, V_C^{nt}, V_C^{ps}, V_C^{pt}, V_{ps-pt}, V_{ns-nt}, V_{fusion}\right) \tag{4}$$

Feed the fusion vector V into the Bi-LSTM model with an attention mechanism. Since our input sequence is short and we use Bi-LSTM, we can overcome the shortcomings of LSTM in other application scenarios. Through the LSTM model, we obtain a feature vector that has been fused by attention.

One kind of recurrent network RNN is called long short-term memory (LSTM). To address the issues of gradient explosion and disappearance in the RNN model during training, a unique gate structure is employed. Figure 1 depicts the LSTM model's basic unit structure. The sigmoid function functions as a gating signal, activating the data and controlling the amount of information that accumulates or is forgotten. The calculation formula of LSTM is as follows:

$$f_t = \sigma\left(w_f \cdot [h_{t-1}, v_t] + b_f\right) \tag{5}$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, v_t] + b_i) \tag{6}$$

$$\widetilde{C_t} = tanh(w_C \cdot [h_{t-1}, v_t] + b_C) \tag{7}$$

$$C_t = f_t * C_{t-1} + i_t * C_t \tag{8}$$

$$O_t = \sigma(w_o \cdot [h_{t-1}, v_t] + b_o) \tag{9}$$

$$h_t = O_t * tanh(C_t) \tag{10}$$

Where $\sigma(\cdot)$ 和 $tanh(\cdot)$ denote activation function. The calculation formula is: $\sigma(x) = \frac{1}{1+e^{-x}}$, $tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, v_t is the input at time-step t, f_t is the forget gate, i_t and O_t denote the input and output gates, C_t is the cell state, h_t is the hidden state.

Bi-directional long short-term memory network (Bi-directional long short-term memory, Bi-LSTM) is an extension of the LSTM model, which includes forward LSTM and backward LSTM. The application of forward and backward LSTM can improve the long-term dependence of model learning. , thereby improving the accuracy of the model. We feed the vector V into the Bi-LSTM model to generate the forward hidden state $\overrightarrow{h_t}$ and backward hidden state $\overleftarrow{h_t}$ of the feature vector V, and splice the forward hidden vector and the backward hidden vector to obtain the final represents the vector h_t.

$$h_t = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right] \tag{11}$$

After the bidirectional LSTM layer, the attention mechanism is applied. The crucial information gleaned from the entire sequence can be improved by adding the Attention mechanism, which can give each feature a distinct weight.

$$u_t = tanh(W_w h_t + b_w) \tag{12}$$

$$a_t = \frac{exp(u_t{}^T u_w)}{\sum_t xp(u_t{}^T u_w)} \tag{13}$$

$$\mathrm{S} = \sum_t a_t h_t \tag{14}$$

The word context vector u_w is randomly initialized and jointly learned during the training process. First, h_t obtains u_t through linear transformation, then multiplies the tyranny of ut with the context vector u_w , and then undergoes softmax normalization to obtain the weight a_t. Finally, let a_t and h_t be multiplied and summed to obtain the weighted backing vector representation S.

D. Classifier

The neural network model extracts more and more features as the number of iterations increases during training, and the training set eventually converges and produces superior results. However, this so-called overfitting problem occurs when testing or validation using this model yields subpar findings. The model's ability to predict unknown samples will be mediocre due to overfitting, and its generalization capacity will be weak. To prevent the network model from overfitting, a Dropout layer and a BatchNormalization layer are added to the model. Dropout can randomly discard some neurons during training to reduce node dependence, but the overall architecture remains unchanged during testing. It is widely used in neural networks and makes the training of deep neural networks possible. Training will cause a shift in the data distribution because the model consists of multiple layers. To keep training stable and increase the training pace, regularization is employed. Using the matching layer, the technique of measuring similarity is chosen. After being connected to BatchNormalization and Dropout, the semantic representation vectors of usernames derived from the attention layer are ultimately connected to the fully connected layer and subsequently to the softmax layer for categorization.

IV. Experiments and Results

A. Dataset

To prove the effectiveness of the algorithm, the Douban-Weibo datasets [33] , WB-BL datasets and WB-DY datasets [17] are used to validate the experiment. Douban–Weibo dataset contains social network topology, user attributes, and user-generated contents. WB-BL and WB-DY contain username and user avatar. The statistics are presented in Table 1.

TABLE I. Statistics of the datasets.

	Dataset		
	Douban-Weibo	***WB-BL***	***WB-DY***
Size	9616	5577	2007

B. Experiment details

Our experimental environment consists of a server with an i9-9900k 3.6GHz CPU and 32GB RAM, equipped with a 2080ti GPU. The operating system is Windows (x64) version 10, and we are using PyTorch 1.13 as the underlying framework. Due to the limited amount of user identity link data, we initialize the model with pre-trained parameters from BERT and CLIP to extract semantic features from both text and images.

Initially, we employ a semantic feature extraction layer to extract the semantic features of text and images. Subsequently, we utilize a bidirectional LSTM with attention mechanism to fuse the feature vectors. In the final layer of the experiment, we use Softmax to calculate the class probabilities. Additionally, the activation function used in the other layers is ReLU.

For the classification model, we utilize a cross-entropy loss function in the experiments, with a Dropout ratio set to 0.5. The experiment is trained for 50 epochs and incorporates early stopping requirements. The batch size is set to 16, and the learning rate for pre-trained layers is 2e-5, while for other layers it is 1e-3. Furthermore, there is a decay of 0.1% during the middle period.

C. Compare with other experiments

To evaluate the performance of our method, we compare it with several state-of-the-art methods listed as follows.

a) Zafarani et al. [21] exploited username redundancy across social networking sites to create a Mobius technique based on naming patterns that worked well across multiple networks.

b) Li et al. [22] analyzed users' unique naming patterns, utilized information redundancy of usernames to construct features, and applied machine learning techniques to identify users using hand-crafted features.

c) Ye et al. [17] apply the multi-lingual BERT pre-training model to extract the deep semantic features of usernames.

d) Shu et al. [13] emphasized the relationship between the user's personality, actions, and behavior and the selection of profile images. The pre-trained VGG16 model is used to classify user avatars.

e) Anisa et al. [28] matched the user profile based on the similarity measure of attribute information such as user name, gender, avatar, etc. in the user profile. Use Levenshtein distance to calculate the user name similarity. Use openface to measure the similarity of user avatars.

Table II depict the results. Inspired by Shu's [13], vgg16 is used to extract user avatar features and identify users across social networks based on user avatar similarity. This shows that the accuracy of using only user avatar features is not as high as that of using only username features. MOBIUS[21], Li's [22] and Ye's [17] utilize username features for user identification across social networks. Compared with the MOBIUS [21] method, our method has improved F1-score by 6.17% on Douban–Weibo, 1.62% on WB-DY, 2.1% on WB-BL. This shows that the method of fusing multi-dimensional user attribute features is more effective than the method based on a single user attribute feature. Both Anisa's [28] and our methods perform user identification across social networks based on multi-dimensional user attributes. Compared with the Anisa's [28] method, our method has improved ACC by 18.1% on Douban–Weibo, 13.7% on WB-DY, 13.8% on WB-BL. The results show that our method can significantly improve the performance of user identification.

D. Ablation experiment

We analyze the contribution of different components by reducing the corresponding modules. The ablation experiment results are shown in Table III. Model 1 uses the BERT model to extract the semantic features of user names as feature representations. Model 2 uses the clip model to extract semantic

features of user avatars as feature representations. Model 3 fuses the semantic features of user names and user avatars as feature representations. Model 4 sends the fused feature vector to LSTM for classification. Model 5 to feed the fused feature vector into the Bi-LSTM model. Model 6 feeds the fused feature vector into the Bi-LSTM model with an attention mechanism. As can be seen from Table III, fusing the semantic features of user names and user avatars can improve the accuracy of user identification.

TABLE II. COMPARISON BETWEEN THE PROPOSED APPROACH AND OTHER MODELS.

Dataset	**Method**	***ACC***	***F1***	***PRE***	***REC***
Douban-Weibo	MOBIUS [21]	0.7875	0.7409	**0.9504**	0.6071
	Li's [22]	0.7916	0.7491	0.9425	0.6216
	Ye's [17]	0.8101	0.7711	0.8828	0.6845
	Shu's [13]	0.6237	0.6538	0.5105	**0.9091**
	Anisa's [28]	0.6925	0.5870	0.8955	0.4366
	Ours	**0.8179**	**0.7866**	0.9149	0.6898
WB-DY	MOBIUS [21]	0.8406	0.8266	0.9502	0.7314
	Li's [22]	0.8443	0.8290	0.9650	0.7266
	Ye's [17]	0.8385	0.8243	0.9683	0.7176
	Shu's [13]	0.6522	0.7177	0.6048	**0.8824**
	Anisa's [28]	0.7559	0.7101	0.9266	0.5755
	Ours	**0.8509**	**0.8400**	**0.9692**	0.7412
WB- BL	MOBIUS [21]	0.8391	0.8171	**0.9559**	0.7135
	Li's [22]	0.8354	0.8140	0.9458	0.7144
	Ye's [17]	0.8434	0.8221	0.9213	0.7421
	Shu's [13]	0.6749	0.6832	0.5888	**0.8136**
	Anisa's [28]	0.7490	0.6878	0.9209	0.5489
	Ours	**0.8523**	**0.8342**	0.9379	0.7511

V. CONCLUSION

In this paper, we propose a multi-modal semantic fusion user identity link classification framework model that combines the CLIP model. This framework utilizes CLIP and Bert pre-training models for semantic feature extraction and integrates multi-modal semantic features from user profiles to enhance the model's effectiveness, thereby achieving improved results on the original semantic features. We conduct comprehensive experiments and comparisons on three datasets using various methods to validate the authenticity and stability of our experiments. In the future, we plan to incorporate additional sequential features, such as user content, to further enhance the accuracy of user identity link classification.

TABLE III. COMPARATIVE RESULTS OF MODEL ABLATION EXPERIMENTS.

Dataset	**No.**	***Method***	***ACC***
Douban-Weibo	1	BERT	0.6202
	2	CLIP	0.6530
	3	BERT + CLIP	0.7034
	4	BERT + CLIP +LSTM	0.7932
	5	BERT + CLIP +BiLSTM	0.8153
	6	BERT + CLIP +Bi-LSTM+Attention	**0.8179**
WB-DY	1	BERT	0.6762
	2	CLIP	0.7173
	3	BERT + CLIP	0.7098
	4	BERT + CLIP +LSTM	0.6770
	5	BERT + CLIP +BiLSTM	0.8261
	6	BERT + CLIP +Bi-LSTM+Attention	**0.8509**
WB- BL	1	BERT	0.6719
	2	CLIP	0.7028
	3	BERT + CLIP	0.7238
	4	BERT + CLIP +LSTM	0.8143
	5	BERT + CLIP +BiLSTM	0.8501
	6	BERT + CLIP +Bi-LSTM+Attention	**0.8523**

ACKNOWLEDGMENT

This research is supported in part by the National Natural Science Foundation of China under Grant No.61672179.

REFERENCES

[1] K. Shu, S. Wang, J. Tang, R. Zafarani, and H. Liu, "User identity linkage across online social networks: A review," ACM SIGKDD Explorations Newsletter, vol. 18, no. 2, pp. 5–17, 2017.

[2] Cao, D., He, X., Nie, L., Wei, X., Hu, X., Wu, S., Chua, T.: Cross-platform app recommendation by jointly modeling ratings and texts. ACM Trans. Inf. Syst. vol.35, no.4, pp. 37:1–37:27, 2017.

[3] R. Zafarani and H. Liu, "Connecting Corresponding Identities across Communities,"Proc. 3rd Int. ICWSM Conf., pp. 354-357, 2009.

[4] A. Agarwal and D. Toshniwal, "Smpft: social media based profile fusion technique for data enrichment,"Computer Networks, vol. 158, pp. 123–131, 2019.

[5] L. Xing, K.K. Deng, H.H. Wu, P. Xie, M.C. Zhang, Q.T. Wu, "Exploiting Two-Level Information Entropy across Social Networks for User Identification."Wireless Communications & Mobile Computing (Online).2021

[6] J. Shu, J. Shi, L. Liao, "Link Prediction Model for Opportunistic Networks Based on Feature Fusion,"IEEE Access, vol. 10, pp.80900–80909, 2022.

[7] T. Ma, L. Guo, Y. Qian, and Y. Tian, "Friend closeness based user matching cross social networks,"Math. Biosci. Eng., vol. 18, no. 4, pp.4264–4292, 2021.

[8] L. Xing, K. Deng, H. Wu, P. Xie, and J. Gao, "Behavioral habits-based user identification across social networks,"Symmetry, vol. 11, no. 9, 1134,2019.

[9] K. Deng, L. Xing, L. Zheng, H. Wu, P. Xie, and F. Gao, “A user identification algorithm based on user behavior analysis in social networks,”IEEE Access, vol. 7, pp. 47114–47123, 2019.

[10] W. Chen, W. Wang, H. Yin, L. Zhao, and X. Zhou. “HFUL: A hybrid framework for user account linkage across location-aware social networks,”VLDB J, vol. 32, pp. 1-22, 2023.

[11] W. He, Y. Li, Y. Zhang and X. Li, “A Binary-Search-Based LocalitySensitive Hashing Method for Cross-Site User Identification,” IEEE Trans. Comput. Soc. Syst., 10(2): pp.480–491, 2022.

[12] R. Zafarani, H. Liu, “Connecting Corresponding Identities across Communities.,” *Proc. 3rd Int. ICWSM Conf.,* pp. 354-357, 2009.

[13] K. Shu, X. Zhou, S. Wang, R. Zafarani, and H. Liu, “The role of user profiles for fake news detection,” in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, Aug. 2019, pp. 436–439

[14] L. Ranaldi and F. M. Zanzotto, "Hiding Your Face Is Not Enough: user identity linkage with image recognition", *Social Network Analysis and Mining*, vol. 10, no. 1, pp. 1-9, 2020.

[15] Y. Huang, P. Zhao, Q. Zhang, L. Xing, H. Wu, and H. Ma. 2023. “A semantic-enhancement-based social network user-alignment algorithm ,” Entropy vol. 25, no. 1, 172 , 2023.

[16] X. Du, S.Y. Chen, Z. Liu, et al., “Multiple userids identification with deep learning,” Expert Syst. Appl., vol. 207, 117924, 2022.

[17] C. Ye, J. Yang, Y. Mao, “User identification for knowledge graph construction across multiple online social networks,” Alexandria Eng. J., vol. 73, pp. 145-158, 2023.

[18] X. Pan, T. Ye, D. Han, S. Song, and G. Huang, “Contrastive languageimage pre-training with knowledge graphs,” in Advances in Neural Information Processing Systems, 2022.

[19] lec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. “Learning transferable visual models from natural language supervision,” In International Conference on Machine Learning, pages 8748–8763. PMLR, 2021. 1, 7, 8, 9

[20] Y. Li, Y. Peng, Z. Zhang, H. Yin, and Q. Xu, “Matching user accounts across social networks based on username and display name,”World Wide Web, vol. 22, no. 3, pp. 1075-1097, 2019.

[21] R. Zafarani, L. Tang, and H. Liu, “User Identification Across Social Media,”ACM Trans. Knowl. Discov. Data, vol. 10, no. 2, pp. 1–30, 2015.

[22] Y. Li, Y. Peng, W. Ji, Z. Zhang, and Q. Xu, “User Identification Based on Display Names Across Online Social Networks,”IEEE Access, vol. 5, pp. 17342-37353, 2017.

[23] Y. Li, H. Cui, H. Liu, X. Li, “Display Name-Based Anchor User Identification across Chinese Social Networks,”In Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 11–14 October 2020; pp. 3984–3989.

[24] Y. Li, and H. Liu. “DENA: display name embedding method for Chinese social network alignment,”Neural. Comput. Appl., vol. 35, no. 10, pp. 7443-7461, 2023.

[25] M. Wang, W. Wang, W. Chen, L. Zhao, “EEUPL: Towards effective and efficient user profile linkage across multiple social platforms,” World Wide Web , vol.24, pp.1731–1748,2021.

[26] L. Wang, K. Hu, Y. Zhang, and S. Cao, “Factor graph model based user profile matching across social networks,” IEEE Access, vol. 7, pp. 152429–152442, 2019.

[27] V. Sharma and C. Dyreson, "Linksocial: linking user profiles across multiple social media platforms", *2018 IEEE International Conference on Big Knowledge (ICBK)*, pp. 260-267, 2018.

[28] A. Halimi and E. Ayday, “Profile matching across online social networks,” *Proc. 22nd International Conference on Information and Communications Security (ICICS)*, pp. 54-70, August 2020.

[29] A. Halimi and E. Ayday. “Efficient quantification of profile matching risk in social networks using belief propagation,” In European Symposium on Research in Computer Security, pages 110--130. Springer, 2020.

[30] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, “Deep contextualized word representations, arXiv preprint arXiv:1802.05365, 2018.

[31] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, “Improving language understanding by generative pretraining,” https://s3-us-west-2. amazonaws.com/openai-assets/research-covers/language-unsupervised/language understanding paper.pdf, 2018.

[32] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, “Bert: pre-training of deep bidirectional transformers for language understanding,” in: Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol., vol. 1, pp. 4171–4186, 2019.

[33] Chen, S.Y.; Wang, J.H.; Du, X.; Hu, Y.Q. A Novel Framework with Information Fusion and Neighborhood Enhancement for User Identity Linkage. In Frontiers in Artificial Intelligence and Applications, Proceedings ofthe 24th European Conference on Artificial Intelligence (ECAI), Online/Santiago de Compostela, Spain, 29 August–8 September 2020; Ios Press: Amsterdam, The Netherlands, 2020; Volume 325, pp. 1754–1761.