Predicting and Analyzing Privacy Settings and Categories for Posts on Social Media

Hsin-Yu Chen
National Cheng Kung University
Tainan, Taiwan
d0107330@gmail.com

Cheng-Te Li
National Cheng Kung University
Tainan, Taiwan
chengte@ncku.edu.tw

2022 IEEE International Conference on Big Data (Big Data) | 978-1-6654-8045-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigData55660.2022.10020677

***Abstract*—While social media is prevalent in people's daily life, privacy control of user-generated posts is becoming increasingly important. In this paper, we propose to enable automatic privacy control for social media posts through two tasks, *predicting privacy settings* and *predicting privacy categories*. The former is to recommend the proper settings of privacy levels, including family, close, casual, and outside, for a post. The latter is to predict the categories of privacy concerns for a post. We propose a multi-task learning-based approach, along with learning feature representation of each post, for such two tasks. Experiments conducted on a real dataset with tweet posts exhibit promising performance of our model, and thus encourage further investigation of privacy-related tasks for privacy control on social media. We also provide a series of extensive analysis with insights that reveal the hidden correlation between privacy settings/categories and post texts.**

***Index Terms*—Privacy Setting Prediction, Privacy Category Prediction, Multi-Task Learning, Embedding Learning, Social Media**

I. Introduction

While interacting with others in social media has become prevalent, many users have become concerned with the privacy of their posts. The default privacy setting of posts is usually to be publicly accessible, and may put personal and sensitive content of users at high risk of privacy leakage. People tend to manually determine the privacy levels of their posts. For example, after preparing a post, one may need to choose its privacy setting from *family*, *close*, *casual*, and *outside*, whose privacy level is from high to low. Nevertheless, more often than not, users are not aware of privacy management and expose their posts under privacy leakage [9], [15]. Therefore, this paper aims to *automatically* recommend the most proper privacy settings for user-generated posts on social media. We believe that an effective and accurate recommendation of privacy settings of posts can not only protect the sensitive content of users from privacy leakage, but also enhance user experience.

In the literatures, limited studies aim at privacy prediction for posts on social media. Mao et al. [12] assume the privacy of posts can be divided into three categories: vacation, drunk, and disease, and make the classification. Caliskan et al. [1] perform binary classification to detect privacy-sensitive posts. Zhang et al. [19] focus only on location privacy of posts. Hsieh et al. [5] predict friendships based on the geographical footprints of users. Lai et al. [11] infer sensitive demographic attributes of users using their interactions with items in the context of recommender systems. Hsieh and Li [6] further highlight private attributes can be predicted through the learning of graph neural networks, and present a defense mechanism to protect user privacy.

In this paper, we make the recommendation of privacy control for posts by performing two tasks. One is *predicting privacy settings*, and the other is *predicting privacy categories*. The first task is to classify a post into four privacy levels, i.e., *family*, *close*, *casual*, and *outside*. The four privay levels mean four tier social circles, *family members*, *close friends*, *casual friends* and *outsider audience* respectively. To be specific, since it is subjective to set the privacy level of a post, we formulate the problem as predicting the probability distribution of four privacy levels for a post. The second task is to classify a post into multiple privacy categories. In the dataset we employed for the experiments, there are 32 privacy categories, such as "relationship status", "religion", "health conditions", "self promotion", "general complaints", and "salary". And each post can belong to multiple categories. Better prediction of privacy categories can help understand how people choose their privacy settings, and benefit privacy management.

Since the privacy categories of a post may be correlated with its privacy settings/levels, we think such two tasks can be mutually reinforced to enhance the prediction performance. Therefore, we propose a multi-task learning-based approach to jointly learn and make predictions for such two tasks. The novelty of this work is three-fold. First, we allow the input features to be shared between two tasks. Second, we perform end-to-end multi-task learning, in which the predicted privacy categories are used to predict the privacy settings. Third, in distilling useful features from posts, we learn the post representation via graph embedding. Experiments conducted on a real dataset show promising performance in both tasks, comparing to several baselines. Extensive studies also deliver several insights revealing the underlying connection between post texts and privacy control.

II. The Proposed Method

The proposed method consists of three parts. We first present the list of extracted features, then describe how to learn the feature representation of each post via graph embedding. Last, we present the proposed multi-task learning model.

A. Feature Extraction

To distill information about privacy from posts, we extract the following privacy-related features.

- *LIWC.* Linguistic Inquiry Word Count (LIWC) is a psycholinguistic lexicon analysis package. LIWC is effective in predicting user personality and sentiments [8], [18]. We think user privacy can be reflected by personality, and thus extract LIWC features. Given a post, LIWC outputs a vector of 70 dimensions[1], corresponding to 70 categories, to represent user personality.
- *Privacy Dictionary.* We use a linguistic-based privacy dictionary [17] that contains an up-to-date list of privacy keywords to distinguish privacy-related words from non-privacy-related ones.
- *Sentiment Detection.* While sentiments had been analyzed to have high correlation to different levels of privacy settings [2], [3], we think detecting post sentiments can benefit the prediction of privacy settings. We use *Stanford NLP sentiment classifier*[2] to detect post sentiments. Each post is assigned to one polarity among *very negative*, *negative*, *neutral*, *positive*, and *very positive*.
- *Sentence Embedding.* The semantics hidden in sentences can also affect the privacy settings of posts. We use *sentence2vec*[3] to extract an embedding vector for each post, due to the nature of short length for social media posts.
- *Meta Information.* We use metadata of each post to be additional features. The metadata includes the presence of hashtags, slang words, images, emojis, and user mentions. The metadata is provided by users to enrich the sentiments and enable social interactions, which may be correlated with privacy settings.

We denote the vector of post i's extracted features as $F^E(i)$.

B. Feature Learning

We consider the underlying correlation between posts to extract interaction-based features for each post. The idea is that posts are correlated with each other in various aspects. Those posts assigned similar privacy settings may be highly correlated with each other. Since each post is composed by words and sentences, the relationships between posts can to some extent reveal their latent privacy correlation. In other words, posts using similar privacy settings or privacy categories could form some relationships. We explore and exploit such latent relationships through feature learning.

To implement such idea, an intuitive solution is to construct a graph $\mathbf{S}$ that captures the correlation between posts, then leverage truncated random walks in graph $\mathbf{S}$ to learn the feature representation of each post [4]. However, this intuitive solution has some disadvantages. First, when the amount of posts increases, it is computationally costly to compute pairwise similarity between posts so that $\mathbf{S}$ can be obtained. Second, graph $\mathbf{S}$ is often quite dense, which makes truncated random walks in graph $\mathbf{S}$ inefficient. Therefore, we extend FeatWalk [7], which is an alternative way to simulate the similarity-based random walks in graph S, to obtain the learned feature of each post. For simplicity, below we use $\mathbf{X}$ to represent the extracted features $F^E(i)$.

Our method consists of three steps. First, we compute the L2 norm to normalize each row of feature matrix $\mathbf{X}$, then remove the values less than zero so that random walks can be performed more efficiently. Each normalized row is denoted as $\bar{x}$. Then we normalize each row of $\mathbf{X}$ again with L1 norm, and a new matrix $\mathbf{Y}$ can be obtained, given by:

$$y_{im} = \frac{\bar{x}_{im}}{\sum_{p=1}^{M} \bar{x}_{ip}},$$

where M is the feature dimension. Second, given a post i, we randomly select a feature column based on its normalized feature, i.e., $\bar{x}_i$. Let a_m denote the m-th feature column in $\mathbf{Y}$. We define the probability of selecting a_m as the next randomly-walked feature:

$$P(i \to a_m) = y_{im}.$$

Features with higher normalized values tend to be selected. Third, given that a_m is selected, we randomly select a post j based on the m-th column of $\mathbf{Y}$. The probability of walking from feature column a_m to post j is defined as:

$$P(a_m \to j) = \frac{y_{im}}{\sum_{n=1}^{N} y_{nm}},$$

where N is the number of posts. Posts possessing higher values on feature m tend to be selected. In such a way, we accomplish the walk from post i to j. Repeating this process multiple times can generate a sequence of posts from every post i, denoted by $Q^{(i)}$, which depicts the trajectory of random walk. Equipped with post sequences, we apply a word embedding method with skip-gram model [13] [14] to learn the feature representation of each post. In natural language processing, the skip-gram architecture learns relations between words and their context. Here each post in feature matrix $\mathbf{X}$ is treated as a word, and random walk paths are considered as sentences. Unless we specify otherwise, we set the length of random walks $l = 35$, and the dimension of features $d = 128$. For each post in $\mathbf{X}$, 25 random walks ($r = 25$) are generated. We denote the learned features of each post i as $F^L(i)$. The extracted and learned feature vectors $F^E(i)$ and $F^L(i)$ are concatenated for predicting privacy settings.

C. Multi-Task Learning

The proposed multi-task learning network architecture is presented in Figure 1. Given the concatenation of extracted and learned feature vectors $F^E(i)$ and $F^L(i)$ as the input, we create a shared hidden layer with dimensionality 128, followed by two branches of task-specific dense layers. Such two branches are specialized in predicting the privacy categories, and predicting the privacy settings. In predicting privacy categories, we utilize *sigmoid* activation function and use *binary cross entropy* to be the loss function, denoted as

[1]http://www.liwc.net/

[2]http://stanfordnlp.github.io/CoreNLP/

[3]https://github.com/klb3713/sentence2vec

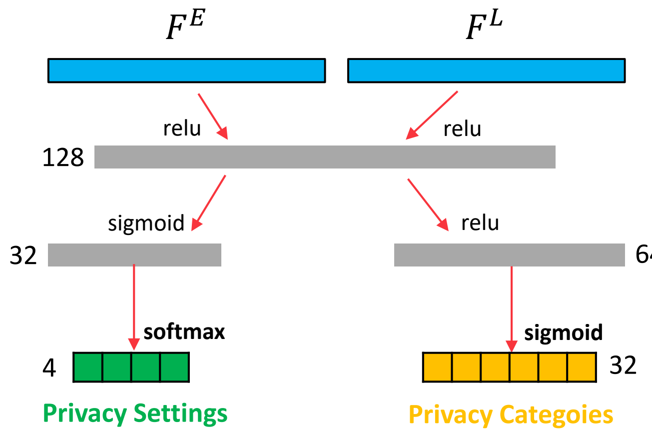

Fig. 1: The proposed network architecture.

TABLE I: MAE results for predicting privacy settings.

	family	close	casual	outside
KNN	0.05393	0.04281	0.03445	0.06392
MTL(FL)	0.05241	0.04194	0.03600	0.06353
RF	0.04880	0.03838	0.03416	0.05904
NN	0.04935	0.03949	0.03399	0.05952
MTL(FE)	0.04673	0.03702	0.03015	0.05718
MTL(FL+FE)	**0.04316**	**0.03327**	**0.02583**	**0.05022**

$\mathcal{L}_c$. In predicting privacy settings, we use *softmax* function to generate the prediction results, i.e., the probability distribution over four privacy settings. We choose *categorical cross entropy* to be the loss function, denoted as $\mathcal{L}_s$. We think that the privacy categories are correlated with privacy settings. That said, posts with similar distributions of privacy categories tend to have similar privacy settings. Therefore, both specific prediction tasks share a hidden layer. Then the shared medium-level features are specialized separately into two tasks. We also use the predicted privacy categories to predict the privacy settings through a hidden-layer combination with the part of privacy setting prediction. The final loss function is designed as:

$$\mathcal{L} = \lambda \cdot \mathcal{L}_c + (1 - \lambda) \cdot \mathcal{L}_s,$$

where $\lambda \in [0, 1]$ balances two tasks. We empirically set $\lambda = 0.9$ using a validation set. We leverage the technique of Adam [10] to optimize the model parameters from training data.

III. Experiments

A. Datasets and Evaluation Settings

Dataset. We use the dataset shared by Song et al. [16]. There are 11,370 tweet posts. The tweet posts had been collected for each privacy category in the pre-defined taxonomy by feeding a list of seed keywords[4] to Twitter Search Service. Each post is associated with 32 privacy categories, which are labeled via Amazon Mechanical Turk (AMT). In addition, each post is also associated with a probability distribution over four privacy settings: *family*, *close*, *casual*, and *outside*. Privacy setting means to share posts with a specific tier of social-circle followers. The above-mentioned four privacy settings represent family members, close friends, casual friends and ousider audience. Each probability represents the percentage of turkers choosing the corresponding privacy setting for a post. Each post can have multiple privacy categories and multiple privacy settings.

In the literature, Caliskan et al. [1] divide social media posts into eights aspects based on human daily behaviors: “location”, “healthcare”, “relationship”, “neutral state”, “emotion”, “activities”, “personal attributes”, and “life milestones.” We follow Song et al. [16] to distribute 32 privacy categories into these eight aspects. That said, we have two granularity levels on privacy categories, and will discuss the prediction performance on both. Figure 2 shows the distribution and the relationships between 32 privacy categories and 8 aspects. It can be clearly found that almost all posts belong to privacy categories of “neutral” and “emotion.” The sample sizes of most categories are quite small, such as “home address”, “salary”, and “status change.” Such an imbalanced distribution makes the prediction of privacy categories challenging.

Baselines. To validate the effectiveness of the proposed multi-task learning model and the feature learning component in predicting privacy settings and categories, we compare the proposed method with three baselines, including Random Forest (RF), K-Nearest Neighbor (KNN) (set $k = 3$), and Deep Neural Network (NN). To have a fair comparison, we construct the same settings as our multi-task learning (MTL) method for NN: set hidden layers (128, 64, 32) in predicting privacy categories, and set hidden layers (128, 32, 4) in predicting privacy settings. These baselines methods use only extracted features (F^E). The proposed MTL model has three variants based on different feature sets, including (a) MTL(FE): using only extracted features (F^E), (b) MTL(FL): using only learned features (F^L), and (c) MTL(FE+FL): using both extracted and learned features.

Evaluation Settings. In predicting the probability distribution of four privacy settings, we choose *mean absolute error* (MAE) as the evaluation metric. For predicting privacy categories. We have two evaluation metrics, *Precision* ($P@K$) and *Recall* ($R@K$). “$P@K$” is the percentage of top K recommended categories that are correctly predicted. “$R@K$” is the percentage of ground-truth categories that are found among top-K predicted categories. While each method generates a probability for each category, we use the prediction probability of each category to rank top K categories. In the settings of training and testing, we conduct 10-fold cross validation to produce the results. The average score of each metric will be presented.

B. Experimental Results

Privacy Settings. The results are shown in Table I. We can observe that the proposed MTL with extracted and learned features outperforms the baselines. Such results demonstrate the effectiveness of multi-task learning. The superiority of

[4]http://sigir16_privacy.farbox.com/

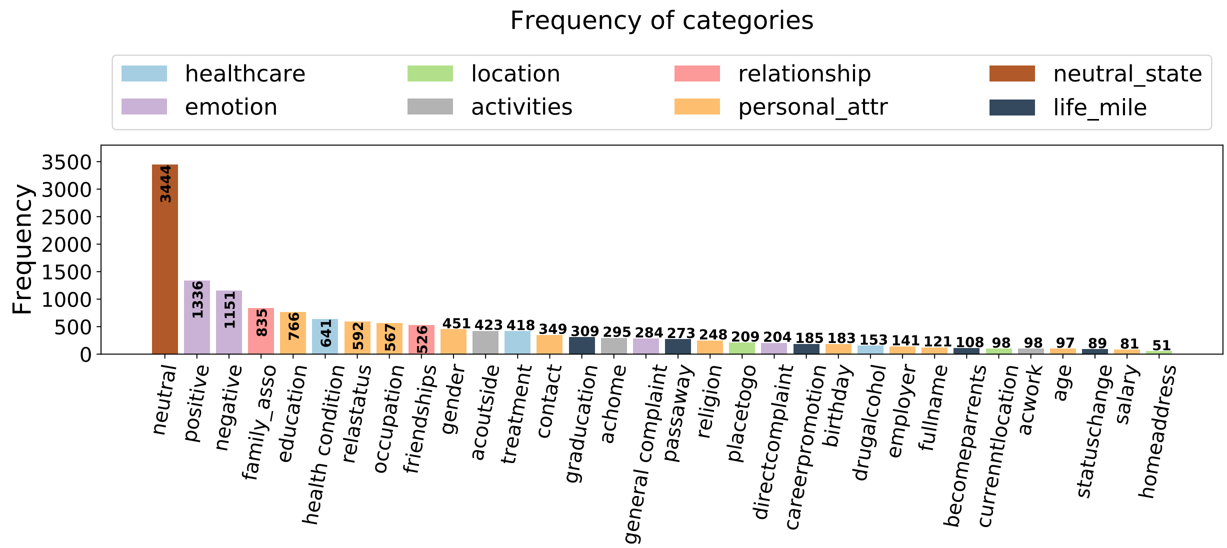

Fig. 2: Distribution of post categories: x-axis displays 32 privacy categories, and y-axis shows the frequency. Eight colors represent 8 privacy aspects.

MTL(FE+FL) is clearer in *family* and *outside* than in *casual* and *close*. We think that the possible reason is the evidences of *casual* and *close* settings in posts and they could be set into *outside* and *family*, respectively. Nevertheless, the positive effect of adding learned features is not so obvious here.

Privacy categories. The results are presented in Table II. The proposed MTL with all features clearly outperforms the baselines. Such results not only validate the usefulness of feature learning, but also prove that there must be some correlation between privacy categories and settings. And multi-task learning is able to effectively learn such correlation for more accurate prediction of privacy categories because it is better than the NN baseline with single objective.

Feature Analysis. We conduct feature analysis to determine which features are important. We employ each of extracted feature sets in our MTL model to see how various features perform. In addition, we use *Extracted Features* that combines all extracted feature sets, *Learned Features* generated from feature learning, and *All* which combined extracted features and learned features. The MAE results in predicting privacy settings are shown in Table III. We can find that *LIWC* and *Sentence Embedding* produce lower MAE values. Such results imply that post semantics represented by *LIWC* and *Sentence Embedding* can provide better predictability and interpretability, especially for "family" and "outside" that may possess stronger clues. Combining all feature sets leads to the lowest MAE values, so every feature set does positively contribute in the prediction. Table IV exhibits the results in predicting privacy categories, which are quite similar to privacy settings, i.e., *LIWC* and *Sentence Embedding* are more effective. Fusing all features can further boost the performance.

Privacy Analysis. We aim to identify the categories that are more challenging to be accurately predicted. The results can be divided into eight privacy aspects and 32 categories, shown in Figure 3 and Figure 4, respectively. Among eight privacy aspects, it can be obviously observed that "relationship", "personal attributes", "neutral state", "emotion", and "healthcare" are better predicted. The possible reasons could be that more posts belong to these aspects, and more distinguishable text evidences (e.g., related words) can be extracted as features for the prediction. Aspects "life milestone", "activities", and "location" can have privacy-related clues belonging to other aspects, and thus are less predictable. The fine-grained results by 32 privacy categories correspond to coarse-grained privacy analysis in colors. We can find that "relationship"-related categories (i.e., *friendships* and *family association*) have higher prediction performance. Similarly, the reason could be they have more posts and more evidential words. In addition, almost half of categories have F1 score close to 0. It should be resulted from few training posts, as pointed out in Figure 2.

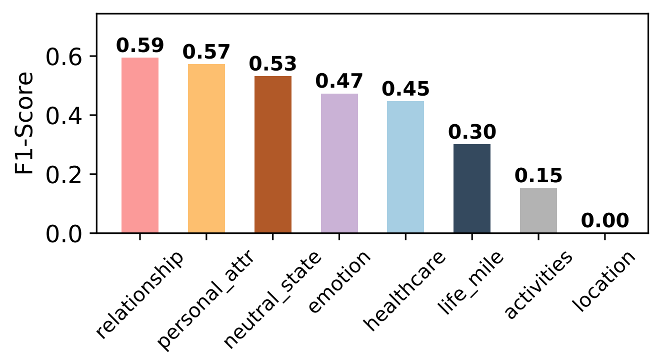

Fig. 3: F1-Score for predicting privacy categories.

Sentiment Analysis. We examine how sentiments over post time affect the prediction performance of privacy categories. There are 5 extracted sentiments: "very negative", "negative",

TABLE II: $P@K$ and $R@K$ for predicting privacy categories.

	P@1	P@3	P@5	R@1	R@3	R@5
KNN	0.2881	0.2392	0.1989	0.2203	0.3635	0.4506
MTL(FL)	0.3670	0.2507	0.2087	0.2391	0.4267	0.4873
RF	0.3712	0.2976	0.2565	0.2755	0.5001	0.5781
NN	0.3925	0.3227	0.2710	0.3004	0.5069	0.5833
MTL(FE)	0.4404	0.3460	0.2848	0.3282	0.5265	0.5950
MTL(FL+FE)	**0.4793**	**0.3781**	**0.3179**	**0.3585**	**0.5784**	**0.6647**

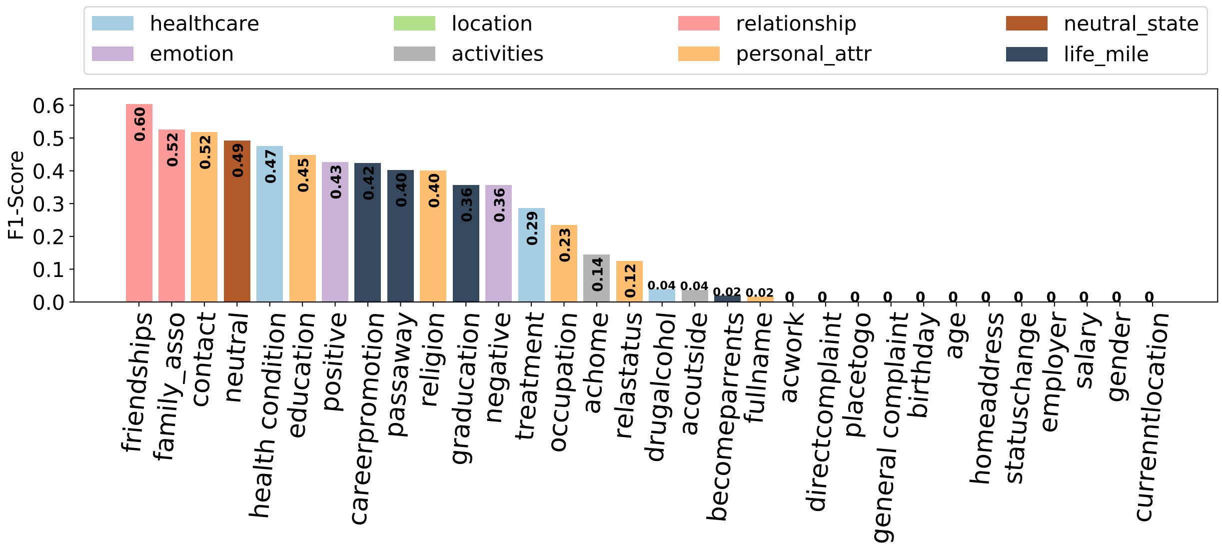

Fig. 4: F1 scores for predicting privacy categories.

TABLE III: MAE for privacy settings with different features.

Features	family	close	casual	outside
LIWC	0.04736	0.03821	0.03048	0.05807
Sentiment	0.05201	0.04139	0.03411	0.06290
Meta Information	0.05198	0.04134	0.03448	0.06282
Privacy Dictionary	0.05219	0.04139	0.03444	0.0628
Sentence Embedding	0.05056	0.03968	0.03365	0.06023
Learned Features	0.05195	0.04145	0.03424	0.06303
Extracted Features	0.04666	0.03758	0.02999	0.05725
All	**0.04373**	**0.03320**	**0.02462**	**0.05181**

"neutral", "positive", and "very positive." The results in Figure 5 show that extreme sentiments at any time have the worst prediction performance due to lacking of training samples. Perhaps it is rare to have extreme sentiments in posts for most people at most time. Negative sentiments tend to have better performance than positive ones by around $4-5\%$. This implies people have high potential to provide clearer text evidences in posts when they feel negative. Furthermore, people tend to post from 5pm to 9pm, i.e., leading to more training posts. Hence, the performance is slightly better in this duration.

We further report the privacy categories and some metadata for posts with good and bad prediction performance, as shown in Table V and Table VI, respectively. Due to page limit, here we only present the top-5 frequent privacy categories. It can be found from Table V that users whose posts belong to positive, negative, neutral, and health condition categories tend to provide stronger text evidences, and thus the privacy of their posts are more predictable. Besides, in Table VI, privacy categories of posts with worse performance are quite diverse, and many of them are unusual (i.e., fewer training samples), e.g., graduation and occupation.

IV. Conclusions

This work aims to predict the privacy settings as well as privacy categories for posts on social media. A better recommendation of privacy control will enable more secure and reliable user experience. We present a multi-task learning-based approach that mutually uses the predictions of privacy categories and privacy settings. We also encode post-post correlation into the feature representation of each post through graph embedding. Experiments conducted on real data demonstrate that our method can significantly outperform baseline models. We believe such results can encourage the research direction of learning-based privacy preservation by incorporating more mutually-correlated tasks.

Acknowledgment

This work is supported by the National Science and Technology Council (NSTC) of Taiwan under grants 110-2221-E-006-136-MY3, 111-2221-E-006-001, and 111-2634-F-002-022.

References

[1] Aylin Caliskan Islam, Jonathan Walsh, and Rachel Greenstadt. Privacy detective: Detecting private information and collective privacy behavior in a large social network. In *WPES*, 2014.

TABLE IV: Results for privacy categories.

Features	P@1	P@3	P@5	R@1	R@3	R@5
LIWC	0.4687	0.3477	0.2875	0.4342	0.5518	0.6283
Sentiment	0.3687	0.2643	0.2161	0.3687	0.4393	0.4930
Meta Information	0.3656	0.2632	0.2185	0.3641	0.4384	0.5026
Privacy Dictionary	0.3969	0.2780	0.2324	0.3850	0.4547	0.5278
Sentence Embedding	0.4031	0.2991	0.2488	0.3913	0.4958	0.5672
Learned Features	0.3656	0.2592	0.2143	0.3656	0.4329	0.4932
Extracted Features	0.4516	0.3330	0.2838	0.4332	0.5542	0.6410
All	**0.4850**	**0.3782**	**0.3167**	**0.4668**	**0.5805**	**0.6685**

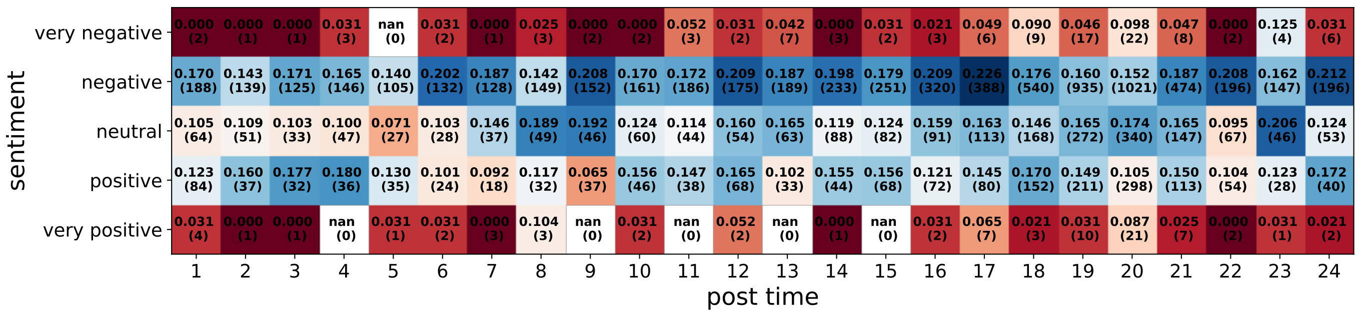

Fig. 5: F1 scores for 5 sentiments over post time in hours. Blue and red indicate better and worse performance, respectively. In each square, the first number is F1, and the number in the parentheses is the sample size.

TABLE V: Privacy categories in good prediction results.

Good Prediction			
sentiment=negative & hour=12 (F1-Score=0.209)		sentiment=negative & hour=17 (F1-Score=0.226)	
aspects	**categories**	**aspects**	**categories**
neutral state	**neutral**	neutral state	**neutral**
relationship	friendships	emotion	**negative**
emotion	**positive**	personal attr	gender
emotion	**negative**	healthcare	**health cond.**
healthcare	**health cond.**	emotion	**positive**

TABLE VI: Privacy categories in bad prediction results.

Bad Prediction			
sentiment=positive & hour=6 (F1-Score=0.101)		sentiment=positive & hour=9 (F1-Score=0.065)	
aspects	**categories**	**aspects**	**categories**
neutral state	**neutral**	neutral state	**neutral**
personal attr	religion	emotion	**positive**
life mile	career promotion	relationship	family asso.
emotion	**positive**	life mile	graduation
healthcare	health cond.	personal attr	occupation

[2] Casey Fiesler, Michaelanne Dye, Jessica L. Feuston, Chaya Hiruncharoenvate, C.J. Hutto, Shannon Morrison, Parisa Khanipour Roshan, Umashanthi Pavalanathan, Amy S. Bruckman, Munmun De Choudhury, and Eric Gilbert. What (or who) is public?: Privacy settings and social media content sharing. In *ACM CSCW*, 2017.

[3] B. Gao, B. Berendt, and J. Vanschoren. Who is more positive in private? analyzing sentiment differences across privacy levels and demographic factors in facebook chats and posts. In *IEEE/ACM ASONAM*, 2015.

[4] Aditya Grover and Jure Leskovec. Node2vec: Scalable feature learning for networks. In *ACM KDD*, 2016.

[5] Hsun-Ping Hsieh, Rui Yan, and Cheng-Te Li. Where you go reveals who you know: Analyzing social ties from millions of footprints. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, pages 1839–1842, 2015.

[6] I-Chung Hsieh and Cheng-Te Li. Netfense: Adversarial defenses against privacy attacks on neural networks for graph data. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2021.

[7] Xiao Huang, Qingquan Song, Fan Yang, and Xia Hu. Large-scale heterogeneous feature embedding. In *AAAI*, 2019.

[8] Clayton J. Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *ICWSM*, 2014.

[9] Jia-Yun Jiang, Cheng-Te Li, and Shou-De Lin. Towards a more reliable privacy-preserving recommender system. *Information Sciences*, 482:248–265, 2019.

[10] Diederick P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

[11] Chih-Te Lai, Cheng-Te Li, and Shou-De Lin. Deep energy factorization model for demographic prediction. *ACM Trans. Intell. Syst. Technol.*, 12(1), nov 2020.

[12] Huina Mao, Xin Shuai, and Apu Kapadia. Loose tweets: An analysis of privacy leaks on twitter. In *WPES*, 2011.

[13] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *NIPS*, 2013.

[14] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *EMNLP*, 2014.

[15] Manya Sleeper, Justin Cranshaw, Patrick Gage Kelley, Blase Ur, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. "i read my twitter the next morning and was astonished": A conversational perspective on twitter regrets. In *ACM CHI*, 2013.

[16] Xuemeng Song, Xiang Wang, Liqiang Nie, Xiangnan He, Zhumin Chen, and Wei Liu. A personal privacy preserving framework: I let you know who can see what. In *ACM SIGIR*, 2018.

[17] Asimina Vasalou, Alastair J. Gill, Fadhila Mazanderani, Chrysanthi Papoutsi, and Adam Joinson. Privacy dictionary: A new resource for the automated content analysis of privacy. *Journal of the American Society for Information Science and Technology*, 62(11):2095–2105, 2011.

[18] Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R. Zaiane. Current state of text sentiment analysis from opinion to emotion mining. *ACM Comput. Surv.*, 50(2):25:1–25:33, 2017.

[19] Yang Zhang, Mathias Humbert, Tahleen Rahman, Cheng-Te Li, Jun Pang, and Michael Backes. Tagvisor: A privacy advisor for sharing hashtags. In *The Web Conference*, 2018.