A Personal Privacy Preserving Framework: I Let You Know Who Can See What

Xuemeng Song[†], Xiang Wang[§], Liqiang Nie[†], Xiangnan He[§], Zhumin Chen[†], Wei Liu[♯]

[†]Shandong University, [§]National University of Singapore, [♯]Tencent AI Lab
{sxmustc, xiangwang1223, nieliqiang, xiangnanhe}@gmail.com, chenzhumin@sdu.edu.cn, wliu@ee.columbia.edu

ABSTRACT

The booming of social networks has given rise to a large volume of user-generated contents (UGCs), most of which are free and publicly available. A lot of users' personal aspects can be extracted from these UGCs to facilitate personalized applications as validated by many previous studies. Despite their value, UGCs can place users at high privacy risks, which thus far remains largely untapped. Privacy is defined as the individual's ability to control what information is disclosed, to whom, when and under what circumstances. As people and information both play significant roles, privacy has been elaborated as a boundary regulation process, where individuals regulate interaction with others by altering the openness degree of themselves to others. In this paper, we aim to reduce users' privacy risks on social networks by answering the question of *Who Can See What*. Towards this goal, we present a novel scheme, comprising of descriptive, predictive and prescriptive components. In particular, we first collect a set of posts and extract a group of privacy-oriented features to describe the posts. We then propose a novel taxonomy-guided multi-task learning model to predict which personal aspects are uncovered by the posts. Lastly, we construct standard guidelines by the user study with 400 users to regularize users' actions for preventing their privacy leakage. Extensive experiments on a real-world dataset well verified our scheme.

CCS CONCEPTS

• **Information systems → Retrieval tasks and goals**; • **Security and privacy → Privacy protections**;

KEYWORDS

Privacy Preserving; Boundary Regulation; Social Media.

1 INTRODUCTION

With the increasing enthusiasm of users to share their daily life on social networks, a large amount of personal data, such as personal demographics, daily activities and even relations with the others, are made publicly available. It is reported that 66% of users' micro-posts are about themselves [24]. The huge amount of users' personal data accessible online may put the users at a high risk of privacy leakage due to the following reasons. On one hand, the default privacy settings usually make UGCs publicly accessible. In fact, people are usually connected with heterogeneous circles on social networks, such as family members, casual friends and even strangers. As a result, UGCs are probably seen by unexpected audience and hence cause unexpected consequences to users. Take a real story as an example. A video podcaster's home was broken into and several video equipments were stolen during his travel. It is ultimately found out that the break-in was caused by his detailed tweets regarding his leave [24]. On the other hand, users may even be unaware of the privacy leakage when they are posting on social networks, which is also the cause of the regrettable messages [36]. Consequently, privacy leakage via user-generated contents (UGCs) in social networks deserves our special attention.

In fact, according to the report [35], 50% of Internet users are concerned about the privacy exposure, up from about 30% in 2009. Privacy is elaborated as a process of boundary regulation [13, 34], where individuals control over how much information about themselves can be divulged to others. Therefore, maintaining appropriate levels of disclosure within one's social environment is of essential significance. In fact, one's social circle can be organized into different groups based on their personal ties with the given user. It is apparent that for different social circles, individuals hold different norms of what kind of information should be treated as privacy. For example, one's age may be kept private to his/her casual friends but visible to family members, while one's negative emotion may be better invisible to family members. Considering that information and audience both play pivotal roles in the privacy preserving, answering the question of *Who Can See What* is essential.

However, answering *Who Can See What* is non-trivial due to the following reasons. First, posts with personal information may explicitly or implicitly convey different aspects of users. These aspects are usually not independent but can be organized into certain structures, such as groups, according to their relatedness. For example, given a set of aspects $\mathcal{I} = \{age, current\ location, places\ planning\ to\ go\}$, aspects "current location" and "places to go" are more correlated and should be modeled together in one group. More often than not, such structure can impose certain constraints to the feature space and enhance the performance of aspect detection. Consequently, the main challenge is how to construct and leverage such structure to

SIGIR '18, July 8–12, 2018, Ann Arbor, MI, USA

ACM ISBN 978-1-4503-5657-2/18/07…$15.00
https://doi.org/10.1145/3209978.3209995

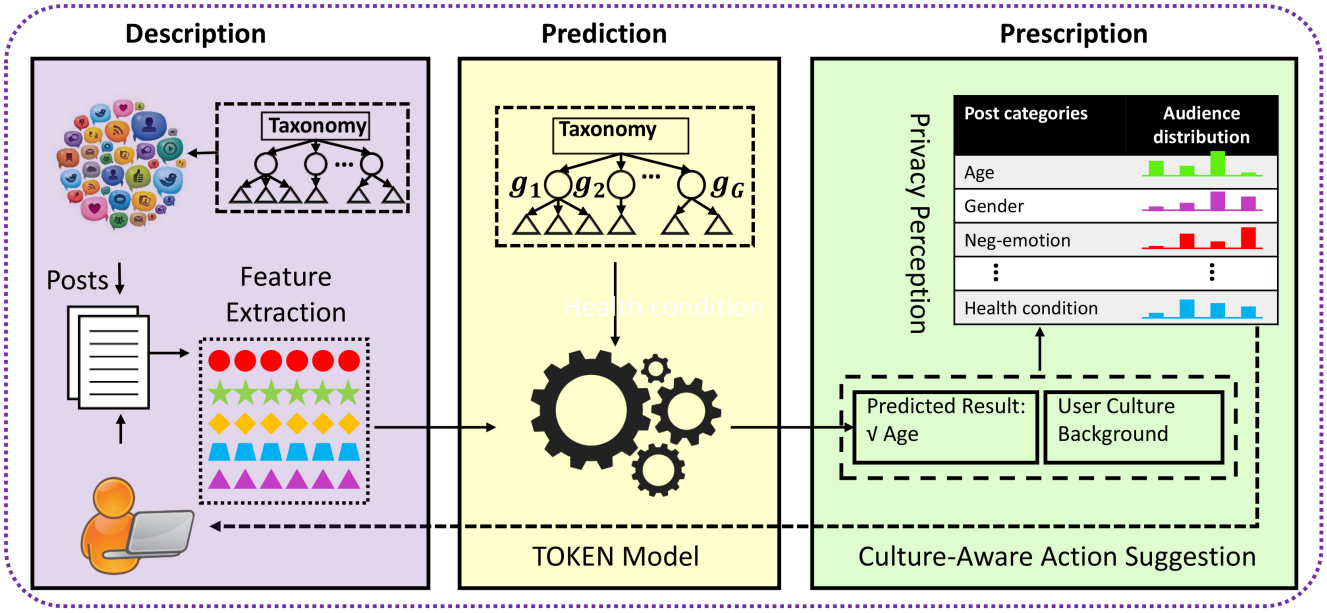


Figure 1: Illustration of the proposed scheme. First, we build a comprehensive taxonomy of the personal aspects, collect a benchmark dataset and extract a rich set of privacy-oriented features from the UGCs. Second, we introduce a taxonomy-constrained model to detect the potential privacy leakage. Last, we suggest users with the possible actions according to the guidelines built via AMT.

learn shared features and specific features. Second, thus far, no gold standard instruction is available to guide *Who Can See What*. As the interpretation of privacy may be subjective and geographically specific, obtaining a unified instruction poses a crucial challenge for us. The third challenge lies in the lack of benchmark dataset and the way to extract a set of privacy-oriented features. This is because it is hard to distinguish the personal posts from the non-personal posts, and some posts are too short to provide sufficient contexts for feature extraction.

To address the aforementioned challenges, we present a novel scheme for boundary regulation, comprising of three components: description, prediction, and prescription. As illustrated in Figure 1, in the first component, we summarize the literature and pre-define a comprehensive taxonomy composed of 32 categories, where each category corresponds to one personal aspect of users. To build a benchmark dataset, we then feed a list of keywords to Twitter Search Service[1] for each category. A set of privacy-oriented features, including linguistic and meta features are extracted to describe the given UGCs. We choose the real-time sharing website Twitter as the study platform due to the following facts: 1) Users in Twitter are keen to share their personal events of various topics; And 2) the followers are broadly and disorderly mixed. Based on these features, the second component then endeavors to discover which personal aspect has been uncovered by the given post. The pre-defined structure in the first component has organized the 32 categories into eight groups, spanning from personal attributes to life milestones. The categories within each group hold both group-sharing features and aspect-specific features. Meanwhile, we assume that there is a low dimensional latent feature space that is capable of capturing the higher-level semantics of UGCs as compared to the original features. To learn the latent feature space and further boost the aspect detection performance, we treat each personal aspect as a task and propose a laTent grOup multi-tasK lEarniNg (TOKEN) model that is able to leverage the pre-defined structure to learn group-sharing latent features and aspect-specific latent features simultaneously. The last component works towards triggering and suggesting users what they should act according to certain guidelines once their privacy leakage is detected by the second component. Considering the existence of cultural difference regarding users' information disclosure norms, we build guidelines by conducting a cross-cultural user study via Amazon Mechanical Turk[2] (AMT). In designing this guideline, we regulate the boundary of users' posts by four tier social circles, namely, *family members*, *close friends*, *casual friends* and *outsider audience*.

Our main contributions can be summarized in threefold:

- We established a taxonomy to comprehensively characterize users' personal aspects. Guided by this taxonomy, we proposed a TOKEN model to uncover the personal aspects disclosed by the user's posts. Regarding the optimization, we theoretically relaxed the non-smooth model to a smooth one and derived its closed-form solution.
- We constructed guidelines regarding users' information disclosure norms with four kinds of social circles. This user study with 400 users cannot be finished without the help of the crowdsourcing Internet marketplace—AMT. In addition, we studied the cultural similarities and differences of users' privacy perception.

[1] https://twitter.com/search-home

[2] https://www.mturk.com/mturk/welcome

- We collected a representative dataset via Twitter Search Service and developed a rich set of privacy-oriented features. We have released the data to facilitate others to repeat experiments and verify their ideas[3].

The remainder of this paper is structured as follows, Section 2 briefly reviews the related work. Sections 3, 4 and 5 present the three components of TOKEN model, namely, description, prediction and prescription, respectively. Section 6 details the experimental results and analyses, followed by our concluding remarks and future work in Section 7.

2 RELATED WORK

Privacy leakage detection and multi-task learning are related to this work.

2.1 Privacy

In the past decades, great efforts have been dedicated to privacy study, and they can be generally divided into two directions. One is investigating privacy issues from the structured data [1, 5, 31], such as users' structured profiles [29, 40], and their privacy settings [21]. Song et al. [40] studied the re-identification problem from users' trajectory records with a human mobility dataset. Besides, Liu et al. [29] proposed a framework for computing privacy scores for users on social networks based on sensitivity and visibility of certain profile items. Han et al. [21] further conducted in-depth study over the privacy issues in people search by simulating different privacy settings in a public social network. In spite of the compelling success achieved by these studies with different application scenarios, far too little attention has been paid to investigate users' unstructured data, whereby the data volume is larger, information is richer, and privacy issues are more prominent, as compared to the structured data.

Another direction is learning privacy issues from the unstructured data [30, 45, 46], mainly referring to UGCs. Approaches following this direction usually focus on training effective classifiers to predict whether the given UGC is privacy-sensitive. Mao et al. [30] studied privacy leakage on Twitter by automatically detecting vacation tweets, drunk tweets, and disease tweets. Caliskan et al. [6] proposed an approach to detecting sensitive content from Twitter users' timelines and associating each user with a privacy score. Although great success has been achieved, they overlooked the relatedness among personal aspects and fed data into traditional machine learning models, such as Naive Bayes [33] and AdaBoost![19]. To bridge this gap, we pre-defined a comprehensive taxonomy to capture users' structural personal aspects and based on which we presented a novel multi-task learning method which considers the relatedness among different personal aspects.

2.2 Multi-task Learning

Multi-task learning works by jointly solving a problem together with other related problems simultaneously, using a shared representation. This often leads to a better model for the research problem, because it allows the learner to use the commonality among the tasks [7]. Hence, precisely identifying and modeling the task relatedness are crucial. Several regularization-style methods have been proposed in the literature to model task relatedness. Argyriou et al. [2] proposed a framework of multi-task feature learning, which learns the shared features among all tasks with convex optimization. The philosophy behind this framework is that all tasks are related, which may be too restrictive and may adversely affect the performance by neglecting the outlier tasks. To address the problem, Song et al. [39] introduced a structure-constrained multi-source multi-task learning model in the context of user interest inference. The authors proposed to construct the task relatedness structure by prior knowledge. However, the underlying assumption of this work that tasks in the same group should share the whole low-level feature space may be unrealistic. Beyond them, we manually pre-defined a taxonomy to structure the task relatedness, and utilized the taxonomy to guide a novel multi-task learning model, which is capable of learning task-sharing and task-specific features. Moreover, we assume that tasks within a group should share certain high-level features.

In fact, MTL has been applied to solve many problems, including social behavior prediction [16], image annotation [14, 15], and web search [3]. However, to the best of our knowledge, limited efforts have been dedicated to applying MTL in the privacy domain, which is the major concern of our work.

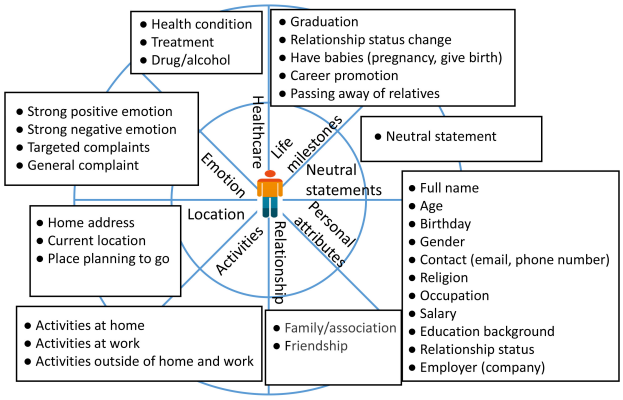


Figure 2: Illustration of our pre-defined taxonomy.

3 DATA AND DESCRIPTION

In this section, we respectively detail the procedure of taxonomy induction, data collection, ground truth construction, as well as feature extraction.

3.1 Taxonomy Induction

In fact, for the privacy detection, Caliskan et al. [6] introduced nine categories: location, medical, drug/alcohol, emotion, personal attacks, stereotyping, family or other associations, personal details, and personally identifiable information. These categories are relatively coarse-grained and hence fail to provide more detailed privacy leakage. In addition, they overlook the life milestones aspect of individuals, which can also be privacy related [12]. Therefore, in this work, by summarizing the literature [6, 12, 25], we pre-defined a comprehensive taxonomy consisting of 32 fine-grained privacy categories. These categories correspond to users' various personal aspects. As shown in Figure 2, these categories can be organized

[3]http://sigir18_privacy.bitcron.com/

into eight groups, namely, *personal attributes*, *relationship*, *activities*, *location*, *emotion*, *healthcare*, *life milestones* and *neutral statements*. Except the *neutral statements* group, categories in the other seven groups are all related to personal issues to some extent. It is noted that, in our work, the neutral statements refer to the posts without revealing any personal information from the other seven groups. Consequently, based on this taxonomy, given a social post, we can group it to at least one category.

3.2 Data Collection

To build our dataset, considering that most of users' private tweets are extremely sparse, we hence did not collect data by following the user-centric policy. Instead, we collected the social posts for each category in the pre-defined taxonomy by respective keywords. In particular, we leveraged the Twitter Search Service. We initially compiled a list of seed keywords[4] for each category and fed them to Twitter Search Service. In the light of this, we obtained 269, 090 raw tweets. To improve the quality of the dataset, we then developed several filter modules for different categories to remove the noise. We filtered out tweets that contain external URLs excluding those referring to users' other social networks' (e.g., Instagram) posts. In addition, as we study the first-order privacy leakage, we ignored retweets in the dataset. Besides, we only retained tweets consisting of more than 50 characters.

3.3 Ground Truth Construction

In our work, we constructed the ground truth about what has been revealed by a given post via AMT. We required workers to annotate each post with multiple categories. It is noted that we only focus on first-order privacy leakage. Particularly, we instructed the AMT workers to annotate a tweet as neutral if it reveals nothing about the tweet owner even it may refer to other people's personal aspects. To ensure the quality of our ground truth, we only employed the AMT masters instead of common workers. AMT masters achieve the "master" distinction by satisfying the demand with a high degree of accuracy. Moreover, we only accepted the submissions whereby the privacy categories labeled by the workers are at least 80% correct based on our sampling validation. To alleviate the problem of subjectivity, we employed three different workers for each post.

At last, we performed the majority voting to establish the final labels for each post and obtained 11, 368 labeled posts. To uncover insights of labeling quality, we use the Fleiss' kappa statistic [18], a variant of Cohen's kappa [44], to measure the inter-worker reliability. Considering that the number of category labels assigned to each tweet is varying, we treat the problem as a set of binary classification. For each binary classification, we count the number of workers who assign this label to the given tweet and those who do not. We finally get the average Fleiss' kappa coefficient as 0.43, which shows a moderate agreement of our workers [28].

3.4 Example Illustration

To get a more intuitive understanding of each category, we take a close look at the samples of each category. Due to the space limitation, we only list a few examples of selected categories in Table 1. We found that users' occupations are mainly revealed by tweeting their new jobs, their feelings about the work, or self-promotion. Users' gender information can be embedded in their roles in relationships (e.g., daughter, wife) or the distinct gender characteristics (e.g., period for women). In addition, users' current locations are usually uncovered by sharing their current feelings or current events. As to general complaints, frequent coughs in the library and unsatisfactory relationships are likely to be complained. Moreover, users are more likely to mention their age when their birthdays are coming. Last but not least, although the neutral statements may talk about "career promotion", "my home" and other personal aspects, they are usually revealing others' privacy or providing no details on the personal information of the user.

Table 1: Examples of selected categories.

Category	Example
Occupation	"I used to be a swimmer... now I'm a coach. And I love torturing my kids."
	"I felt more control of my work as a Teacher. "
Gender	"I seriously going to buy tacos... I am my father's daughter. "
	"The worst thing you do is piss me off while I'm on my period."
Current location	"At the Bell Performing Arts Centre for the LTS Jazz Band Concert #sweet"
	"She told the doctor tomorrow is my birthday I can't be in the hospital"
General complaint	"dude if you're going to cough every 20 seconds in the library can u leave"
	"being in a relationship is stressful i wanna take a nap"
Age	"...when I told him I'm only 24"
	"Hey @user1 its my birthday tomorrow. I am turning 12! "
Neutral statement	"Chelsea look like they got promoted last season.."
	"Do you want my home address and social security too?"

3.5 Features

To capture the user's personal leakage, we extracted a rich set of privacy-oriented features.

3.5.1 LIWC. LIWC, short for Linguistic Inquiry Word Count, is a psycholinguistic transparent lexicon analysis tool, and its effectiveness has been extensively validated in users' personality prediction [37, 38]. Considering that users' personality traits significantly affect their behaviors, including privacy perceptions [26], we adopted the LIWC feature to capture the sensitivity of a given UGC. The main component of LIWC is a dictionary, containing the mappings from words to 70 categories[5]. Given a document, LIWC generates a vector to represent the percentage of words falling into each category. Moreover, we noticed that the 70 categories in LIWC dictionary, such as "job" and "home", can comprehensively cover the user's personal aspects.

3.5.2 Privacy Dictionary. The privacy dictionary [43] is a new linguistic resource for automated content analysis on privacy related texts. We believe that sensitive UGCs should contain some typical privacy related keywords. We hence employed this dictionary to discriminate sensitive and non-sensitive UGCs. This dictionary consists of eight categories[6], derived from a wide range of privacy-sensitive empirical materials. With the help of this dictionary, we can generate similar outputs as LIWC does.

3.5.3 Sentiment Analysis. Different personal aspects are frequently conveyed different sentiments. For example, we observed

[4]These keywords for each category can be available via http://sigir16_privacy.farbox.com/.

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

[6]They are Law, OpenVisible, OutcomeState, NormsRequisites, Restriction, NegativePrivacy, Intimacy, and PrivateSecret.

that people usually broadcast their graduation and becoming parents in a more positive way, while describe their medical treatments in a more negative way. Inspired by this, we adopted the sentiment features [17] and utilized the *Stanford NLP sentiment classifer*[7] to judge tweets' polarity. In particular, we assigned each tweet with a value ranging from 0 to 4, corresponding to *very negative*, *negative*, *neutral*, *positive*, *very positive*.

3.5.4 Sentence2Vector. Considering the short-length nature of tweet, to perform content analysis, we employed the state-of-the-art textual feature extraction tool Sentence2Vector[8]. Sentence2Vector is developed based on the word embedding algorithm Word2Vector [32], which has been found to be effective to alleviate the semantic problems of word sparseness [20]. Given a UGC, Word2Vector would project it to a fixed dimensional space, where similar words are encoded spatially. In our work, we treated each tweet as a sentence, and utilized the Sentence2Vector tool to generate the vector representation of each tweet.

3.5.5 Meta-features. Apart from the above linguistic features, we extracted several metadata features, which have also been verified to be effective in topic detection [41]. These features include the presence of hashtags[9], slang words, images, emojis[10], and user mentions[11]. In particular, to count the number of slang words, we constructed a local slang dictionary, consisting of 5,374 words by crawling the Internet Slang Dictionary & Translator[12]. Moreover, we also incorporated the timestamp as an important feature, as we observed that users would post activities at work in the daytime while post their drug/alcohol aspects in the evening. Notably, we only utilize the post time at the level of hours.

4 PREDICTION

In this section, we detail the prediction component.

4.1 Notation

We first declare some notations. In particular, we use bold capital letters (e.g., $\mathbf{X}$) and bold lowercase letters (e.g., $\mathbf{x}$) to denote matrices and vectors, respectively. We employ non-bold letters (e.g., x) to represent scalars, and Greek letters (e.g., β) as parameters. If not clarified, all vectors are in column form.

In our work, each task is aligned with one personal aspect, and we hence have $Q = 32$ tasks, which have been pre-organized into $G = 8$ groups, according to the pre-defined taxonomy. Meanwhile, we are given N users and each is represented by a D-dimensional vector. Let $\mathbf{X} \in \mathbb{R}^{N\times D}$ stand for the input matrix and $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_Q\} \in \mathbb{R}^{N\times Q}$ denote the corresponding label matrix, where $\mathbf{y}_q = \{y_1, y_2, \cdots, y_N\}^T \in \{1, -1\}^N$ corresponds to the label vector for the q-th task.

[7] http://stanfordnlp.github.io/CoreNLP/
[8] https://github.com/klb3713/sentence2vec
[9] A hashtag refers to a specially designated word prefixed with a '#', which usually represents the topic of this tweet.
[10] An emoji refers to a "picture character" to express facial expressions, concepts, activities and so on.
[11] A user mention is a specially designated word in a tweet, prefixed with a "@", which usually refers to other users.
[12] http://www.noslang.com/

4.2 Model Formulations

For each task, we can learn a predictive model, which is defined as follows,

$$\mathbf{f}_q(\mathbf{X}) = \mathbf{X}\mathbf{w}_q, \tag{1}$$

where $\mathbf{w}_q = (w_q^1, w_q^2, \cdots, w_q^D)^T \in \mathbb{R}^D$ represents the linear mapping function for the q-th task. Let $\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_Q\} \in \mathbb{R}^{D\times Q}$. We adopt the least square loss function to measure the errors,

$$L(\mathbf{W}) = \frac{1}{2N}\left\|\mathbf{Y} - \mathbf{X}\mathbf{W}\right\|_F^2, \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of matrix. $l_{2,1}$-norm has been proven to be effective to select the relevant features for at least one task. In particular, the multi-task learning with $l_{2,1}$-norm is defined as follows,

$$\Gamma = L(\mathbf{W}) + \frac{\beta}{2}\left\|\mathbf{W}\right\|_{2,1}, \tag{3}$$

where β is a non-negative regularization parameter, $\left\|\mathbf{W}\right\|_{2,1} = \sum_{d=1}^{D}\left\|\mathbf{w}^d\right\|$ is the $l_{2,1}$-norm of $\mathbf{W}$, $\mathbf{w}^d = (w_1^d, w_2^d, \cdots, w_Q^d)$, and $\left\|\mathbf{w}^d\right\|$ represents the Euclidean norm of vector $\mathbf{w}^d$. The hidden assumption behind $l_{2,1}$-norm is that all tasks are related and share the common set of relevant features. However, such assumption is not realistic and makes the multi-task learning not robust to the outlier tasks. Beyond that, as aforementioned, all the tasks in our work have been pre-organized into eight groups according to the proposed taxonomy. It is thus reasonable to assume that tasks belonging to the same group would be more likely to share a common set of relevant features. For example, tasks "places planning to go" and "current location" belonging to the location group of the taxonomy may share a common set of location-relevant features. Let C_g stand for the index set of tasks part of the g-th group and the diagonal matrix $\mathbf{V}_g \in \mathbb{R}^{Q\times Q}$ denote the corresponding group assignment. $\mathbf{V}_g(q, q) = 1$ if $q \in C_g$, and 0 otherwise. Thereafter, the objective function in Eqn.(3) can be strengthened as,

$$\Gamma = L(\mathbf{W}) + \frac{\beta}{2}\sum_{g=1}^{G}\sum_{d=1}^{D}\left\|(\mathbf{W}\mathbf{V}_g)^d\right\|. \tag{4}$$

It is worth noting there are two special cases. When the number of groups $G = 1$, where all tasks are learned jointly in one group, it reduces to the traditional multi-task feature learning [2]. When $G = Q$, where all tasks are learned separately, it reduces to the traditional supervised machine learning. Besides, we also argue that tasks of the same group in the taxonomy may not share the common set of low-level relevant features but the common set of high-level latent features. We assume that there are J, where $J \le D$, latent features. Each task is defined as a linear combination of a subset of these latent features. Formally, let us define $\mathbf{W} = \mathbf{L}\mathbf{S}$, where $\mathbf{L} \in \mathbb{R}^{D\times J}$ and $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_Q\} \in \mathbb{R}^{J\times Q}$. Each column of $\mathbf{L}$ stands for a latent feature, and each row of $\mathbf{S}$ represents the linear weights of latent features. We hence impose the $l_{2,1}$-norm on $\mathbf{S}$ instead of $\mathbf{W}$ to learn the group-sharing latent features. Apart from the group-sharing latent features, we also assume each task should be related to a few specific latent features, which is implemented

by the l_1 norm of $\mathbf{S}$. Putting them together, we have the following objective function Γ,

$$L(\mathbf{L},\mathbf{S}) + \frac{\beta}{2}\sum_{g=1}^{G}\sum_{j=1}^{J}\left\|(\mathbf{SV}_g)^j\right\| + \frac{\gamma}{2}\left\|\mathbf{S}\right\|_1 + \frac{\mu}{2}\left\|\mathbf{L}\right\|_F^2, \tag{5}$$

where $\left\|\mathbf{S}\right\|_1$ is the entry-wise l_1 norm of matrix $\mathbf{S}$, while μ and γ are non-negative regularization parameters.

4.3 Optimization

We adopt the alternative optimization strategy to solve $\mathbf{S}$ and $\mathbf{L}$. In particular, we optimize one variable while fixing the other in each iteration. We keep this iterative procedure until the objective function converges.

4.3.1 Computing $\mathbf{L}$ *with* $\mathbf{S}$ *fixed.* We first fix $\mathbf{S}$ and take derivative of the objective function with respect to $\mathbf{L}$. We thus have,

$$\frac{1}{N}\mathbf{X}^T\mathbf{XLSS}^T + \mu\mathbf{L} = \frac{1}{N}\mathbf{X}^T\mathbf{YS}^T. \tag{6}$$

Inspired by the Lemma 4.3.1 in [23], we transform the above equation to the following linear system,

$$\begin{cases} \mathbf{AVec(L)} = \mathbf{B}, \\ \mathbf{A} = [\frac{1}{N}\mathbf{SS}^T \otimes \mathbf{X}^T\mathbf{X} + \mu\mathbf{I}], \\ \mathbf{B} = \mathbf{Vec}(\frac{1}{N}\mathbf{X}^T\mathbf{YS}^T), \end{cases} \tag{7}$$

where $\otimes$ denotes the Kronecker product, $\mathbf{I} \in \mathbb{R}^{(D\times J)\times(D\times J)}$ is an identity matrix, and $\mathbf{Vec}(\cdot)$ stands for stacking columns of a matrix into a single column vector. It is easy to prove that $\mathbf{A}$ is always positive definite [23] and invertible.

4.3.2 Computing $\mathbf{S}$ *with* $\mathbf{L}$ *fixed.* Fixing $\mathbf{L}$ to optimize $\mathbf{S}$, we encounter two non-smooth terms, $l_{2,1}$-norm and l_1 norm, which are intractable to be solved directly. To convert the $l_{2,1}$-norm, we resort to another variational formulation [2, 39] of the $l_{2,1}$-norm in Eqn.(5) as follows,

$$\Gamma = L(\mathbf{L},\mathbf{S}) + \frac{\beta}{2}\Big(\sum_{g=1}^{G}\sum_{j=1}^{J}\left\|(\mathbf{SV}_g)^j\right\|\Big)^2 + \frac{\gamma}{2}\left\|\mathbf{S}\right\|_1. \tag{8}$$

According to the Cauchy-Schwarz inequality, given an arbitrary vector $\mathbf{b} \in \mathbb{R}^M$ such that $\mathbf{b} \neq \mathbf{0}$, we have,

$$\begin{aligned}\sum_{i=1}^{M}|b_i| &= \sum_{i=1}^{M}\theta_i^{\frac{1}{2}}\theta_i^{-\frac{1}{2}}|b_i| \\ &\leq \Big(\sum_{i=1}^{M}\theta_i\Big)^{\frac{1}{2}}\Big(\sum_{i=1}^{M}\theta_i^{-1}b_i^2\Big)^{\frac{1}{2}} = \Big(\sum_{i=1}^{M}\theta_i^{-1}b_i^2\Big)^{\frac{1}{2}},\end{aligned} \tag{9}$$

where θ_i's are introduced variables that should satisfy $\sum_{i=1}^{M}\theta_i = 1$, and $\theta_i > 0$. The equality holds for $\theta_i = |b_i|/\left\|\mathbf{b}\right\|_1$. Based on this, we derive the following inequality,

$$\begin{aligned}\Big(\sum_{g=1}^{G}\sum_{j=1}^{J}\left\|(\mathbf{SV}_g)^j\right\|\Big)^2 &\leq \sum_{g=1}^{G}\frac{\Big(\sum_{j=1}^{J}\left\|(\mathbf{SV}_g)^j\right\|\Big)^2}{\theta_k} \\ &\leq \sum_{g=1}^{G}\sum_{j=1}^{J}\frac{\left\|(\mathbf{SV}_g)^j\right\|^2}{\theta_{k,g}},\end{aligned} \tag{10}$$

where we introduce the variable $\theta_{k,g}$. The equality can be attained if $\theta_{k,g}$ satisfies that,

$$\theta_{k,g} = \frac{\left\|(\mathbf{SV}_g)^j\right\|}{\sum_{g=1}^{G}\sum_{j=1}^{J}\left\|(\mathbf{SV}_g)^j\right\|}. \tag{11}$$

Consequently, fixing $\mathbf{L}$ and minimizing Γ is equivalent to minimizing the following convex objective function,

$$\Gamma = L(\mathbf{L},\mathbf{S}) + \frac{\beta}{2}\sum_{q=1}^{Q}\sum_{j=1}^{J}\frac{\left\|(\mathbf{SV}_g)^j\right\|^2}{\theta_{k,g}} + \frac{\gamma}{2}\left\|\mathbf{S}\right\|_1. \tag{12}$$

To facilitate the computation of the derivative of objective function Γ with respect to $\mathbf{S}$, we define a diagonal matrix $\mathbf{\Theta}_g \in \mathbb{R}^{J\times J}$ as follows,

$$\mathbf{\Theta}_g(j,j) = \frac{1}{\theta_{j,g}}. \tag{13}$$

The final objective function Γ can be rewritten as follows,

$$\Gamma = L(\mathbf{X},\mathbf{Y}) + \frac{\beta}{2}\sum_{g=1}^{G}tr\Big[(\mathbf{SV}_g)^T\mathbf{\Theta}_g\mathbf{SV}_g\Big] + \frac{\gamma}{2}\left\|\mathbf{S}\right\|_1. \tag{14}$$

where $tr(\mathbf{A})$ is the trace of matrix $\mathbf{A}$. To optimize the l_1 norm, we use the fast iterative shrinkage-thresholding algorithm (FISTA) [4] as follows,

$$\begin{cases}\Gamma_q &= h(\mathbf{s}_q) + p(\mathbf{s}_q), \\ h(\mathbf{s}_q) &= L(\mathbf{L},\mathbf{s}_q) + \frac{\beta}{2}\sum_{q\in C_g}tr(\mathbf{s}_q^T\mathbf{\Theta}_g\mathbf{s}_q), \\ p(\mathbf{s}_q) &= \frac{\gamma}{2}\left\|\mathbf{s}_q\right\|_1.\end{cases} \tag{15}$$

The key iteration step of FISTA is to calculate $s_q^{(k)}$ by minimizing the following function,

$$\arg\min_{S}\Big\{p(\mathbf{s}_q) + \frac{R_q^{(k)}}{2}\Big\|\mathbf{s}_q - (\mathbf{z}_q^{(k)} - \tfrac{1}{R_{kq}}\nabla h(\mathbf{z}_q^{(k)})\Big\|_F^2\Big\}, \tag{16}$$

where $R_q^{(k)}$ is the Lipschitz constant of $\nabla h(\mathbf{s}_q)$, $\mathbf{z}_q^{(k)}$ is a linear combination of $\mathbf{s}_q^{(k-1)}$ and $\mathbf{s}_q^{(k-2)}$, and $\nabla h(\mathbf{s}_q)$ is,

$$\nabla h(\mathbf{s}_q) = \frac{1}{N}\mathbf{L}^T\mathbf{X}^T(\mathbf{XLs}_q - \mathbf{y}_q) + \beta\sum_{q\in C_g}\mathbf{\Theta}_g\mathbf{s}_q. \tag{17}$$

We solve Eqn.(16) by the following soft-threshold step,

$$\mathbf{s}_q^{(k)} = \mathcal{T}_{\frac{\gamma}{2R_q^{(k)}}}(\mathbf{e}_q) = max(0, 1 - \frac{\gamma/2R_q^{(k)}}{\left\|\mathbf{e}_q\right\|_1})\mathbf{e}_q, \tag{18}$$

where $\mathcal{T}$ is a shrinkage operator [4] and $\mathbf{e}_q$ is defined as,

$$\mathbf{e}_q = \mathbf{z}_q^{(k)} - \frac{1}{R_q^{(k)}}\nabla h(\mathbf{z}_q^{(k)}). \tag{19}$$

Based on the sub-multiplicative property of spectral norm, we easily derive that $\left\|\nabla h(\mathbf{s}_{q_1}) - \nabla h(\mathbf{s}_{q_2})\right\|$ equals to,

$$\begin{aligned}&\Big\|\beta\sum_{q\in C_g}\mathbf{\Theta}_g(\mathbf{s}_{q_1} - \mathbf{s}_{q_2}) + \tfrac{1}{N}\mathbf{L}^T\mathbf{X}^T\mathbf{XL}(\mathbf{s}_{q_1} - \mathbf{s}_{q_2})\Big\| \\ &\leq \Big(\beta\sum_{q\in C_g}\left\|\mathbf{\Theta}_g\right\| + \frac{1}{N}\left\|\mathbf{L}^T\mathbf{X}^T\mathbf{XL}\right\|\Big)\left\|\mathbf{s}_{q_1} - \mathbf{s}_{q_2}\right\| \\ &\leq R_q\left\|\mathbf{s}_{q_1} - \mathbf{s}_{q_2}\right\|,\end{aligned} \tag{20}$$

Table 2: Performance comparisons of our TOKEN model trained with different feature configurations(%).

Features	S@K			P@K			p-value
	S@1	*S*@3	*S*@5	*P*@1	*P*@3	*P*@5	
Privacy dictionary	8.56 ± 0.73	18.38 ± 0.78	54.26 ± 1.54	8.56 ± 0.73	6.33 ± 0.25	11.28 ± 0.36	5.9e−22
Sentiment	30.48 ± 1.51	52.23 ± 1.09	63.10 ± 1.28	30.48 ± 1.51	17.44 ± 0.36	13.32 ± 0.25	1.6e−20
Meta-features	30.31 ± 1.48	52.28 ± 1.08	63.12 ± 1.23	30.31 ± 1.48	17.38 ± 0.49	13.10 ± 0.65	9.9e−21
Sentence2Vector	33.29 ± 1.77	59.06 ± 0.97	70.91 ± 0.54	33.29 ± 1.77	20.66 ± 0.34	15.54 ± 0.17	2.0e−21
LIWC	37.13 ± 2.45	67.98 ± 1.50	78.65 ± 1.42	37.13 ± 2.45	24.72 ± 0.70	17.44 ± 0.54	3.1e−10
Total	44.37 ± 1.33	74.67 ± 1.38	84.66 ± 0.59	44.37 ± 1.33	28.42 ± 0.57	19.86 ± 0.32	-

whereby we enforce $R_q^{(1)} = R_q^{(2)} = \cdots = R_q$, and $\|\cdot\|$ denotes the spectral norm of matrix as well of Euclidean norm of vector. As $\mathbf{\Theta}_g$ and $\mathbf{L}^T\mathbf{X}^T\mathbf{X}\mathbf{L}$ are both positive-semidefinite matrices, simple algebra computation gives that,

$$R_q = \beta \sum_{q \in C_g} \lambda_{max}(\mathbf{\Theta}_g) + \frac{1}{N}\lambda_{max}(\mathbf{L}^T\mathbf{X}^T\mathbf{X}\mathbf{L}), \tag{21}$$

where $\lambda_{max}(\cdot)$ denotes the maximum eigenvalue of a matrix.

5 PRESCRIPTION

The section details how to construct the guidelines and use the guidelines to recommend appropriate actions to users.

5.1 Guideline Construction

Although privacy may be subjective, there is still a societal consensus that certain information is more private than the others from a general societal view [6]. We thus conducted a user study via AMT to build guidelines regrading disclosure norms in different circles. Considering the existence of cultural difference, we launched a cross-cultural study within two distinct areas: the U.S. and Asia[13], where for each area, we hired 200 subjects. Each subject was required to answer a questionnaire, which consists of a series of questions of whether he/she feels comfortable to share the given personal aspect to four social circles: *Family members*, *Close Friends*, *Casual Friends* and *Outsider Audience*. Finally, we harvested two tables of guidelines, reflecting the privacy perception of users from the U.S. and Asia, respectively.

5.2 Action Suggestion

Based on our proposed TOKEN model, we can infer which personal aspect has been leaked from the given UGC. Once the privacy leakage is detected, assuming that we know the culture background of this user, we then can resort to our culture-aware guidelines established by hundreds of users and choose who is appropriate to see this UGC to avoid the privacy leakage. In a sense, we can remind users of what has been uncovered and accordingly recommend the appropriate UGC-level privacy settings for their social platforms. For example, an Asia girl is posting a tweet pertaining to her health condition, and our scheme would recommend her to set this UGC to be accessed only by her family members and close friends, according to the guidelines.

[13]We found that 99% involved Asia subjects are Indians.

6 EXPERIMENTS

In this section, we conducted extensive experiments to verify the effectiveness of our proposed scheme.

6.1 Experimental Setting

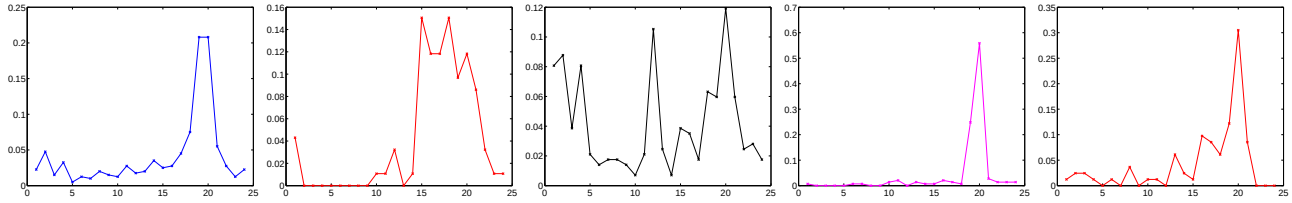
For the task of privacy leakage detection, precision is more important than recall. We hence measured the proposed TOKEN model and its competitors via two widely-used metrics: *S@K* and *P@K* [9, 10, 39]. *S@K* represents the mean probability that a correct privacy category is captured within the top K recommended categories. *P@K* stands for the proportion of the top K recommended categories that are correct. We employed the grid search strategy to obtain the optimal regularization parameters among the values $\{10^r : r \in \{-8, -7, \cdots, 2, 3\}\}$ regarding *P*@1. Experimental results reported in this paper are the average values over 10-fold cross validation.

6.2 Evaluation of Description

To examine the discriminative features we extracted, we conducted experiments over different kinds of features using **TOKEN**. In particular, we also performed significant tests to validate the effectiveness of all the features regarding *S*@5. Table 2 comparatively shows the performance of **TOKEN** in terms of different feature configurations.

It can be seen that our model based on LIWC features achieves the best performance, while the features extracted based on the privacy dictionary are the least effective. This shows that the users' privacy is better characterized by the LIWC dictionary, as compared to the privacy dictionary. One possible explanation is that the categories of LIWC dictionary, include not only content categories such as "home", "job", and "social" that intuitively capture usersą́ŕ personal aspects, but also certain style (function) categories like "pronouns" and "verb tense" that provide the self- or other-references and temporal hints. Meanwhile, although the privacy dictionary is not much powerful when $K = 1$ and $K = 3$, its performance is largely improved when K reaches 5.

Although meta-features only account for six dimensions and the sentiment features are only one-dimensional, they also yield compelling performance. In particular, we argue that the timestamps of UGCs may play an important role regarding privacy leakage detection. We thus had a closer look at the comparison among the time distributions of several representative categories in Figure 3. As can be seen from Figures 3(a), 3(b), and 3(c), categories related to activities show prominent temporal patterns. For example, tweets related to users' activities at home peak at around 12pm and 20pm,



(a) Activities Outside (b) Activities at Work (c) Activities at Home (d) Drug/Alcohol (e) Status Change

Figure 3: Illustration of temporal patterns regarding personal aspects. X axis: the timeline (by hour); Y axis: the distribution of tweets.

while those related to users' activities outside of home and work are more likely to be posted at around 20pm. In addition, Figure 3(d) shows that users would tweet their drug/alcohol aspects in the evening. To some extent, this also reflects the fact that users are more likely to get drunk during their activities outside. Interestingly, we also observed that users prefer to post their "status change" in the evening.

6.3 Evaluation of Prediction

To verify the effectiveness of our proposed model, we compared **TOKEN** with the following four baselines.

SVM: The first baseline is the support vector machine (SVM) [11], which simply concatenates the features generated from different sources into a single feature vector and learns each task individually. We chose the learning formulation with the kernel of radial-basis function. We implemented this method with the help of LIBSVM [8].

MTL_Lasso: The second baseline is the multi-task learning with Lasso [42]. This model also does not take advantage of prior knowledge about tasks relatedness.

MTFL: The third baseline is the multi-task feature learning [2], which takes advantage of the group lasso to jointly learn features for different tasks. However, this model assumes that all tasks are relevant and organizes all tasks in a single group.

GO-MTL(without taxonomy): The fourth baseline is the grouping and overlap in multi-task learning proposed in [27]. This model does not leverage the prior knowledge of task relations, as there is no taxonomy constructed to guide the learning. It is worth mentioning that we can derive **GO-MTL** from **TOKEN** by making $\beta = 0$.

For each method mentioned above, the involved parameters were carefully tuned, and the parameters with the best performance in $S@5$ were used to report the final comparison results. Table 3 shows the performance comparison between the baselines and our proposed **TOKEN**. First, we noticed that **TOKEN** outperforms the single task learning **SVM**. This verifies that there are relationships among tasks. This also shows the superiority of our work over other similar privacy detection researches [6, 30]. In particular, it is not surprised that **SVM** achieves the worst performance. This may be due to insufficient positive training samples for certain categories. For example, there are only 52 positive training samples available for category "home address". Multi-task learning is able to alleviate the unbalanced training sample problems by borrowing some samples from related tasks. In addition, **TOKEN** shows superiority over **MTL_Lasso** and **MTFL**, respectively, which enables us to draw a conclusion that it is effective to learn tasks by groups, defined by the taxonomy. Besides, the less satisfactory performance of **GO-MTL**, as compared to **TOKEN**, also demonstrates the importance to incorporate the prior grouping knowledge of tasks. Moreover, we also performed significant tests over the 10-fold cross validation and found that **TOKEN** can significantly outperforms the baselines regarding $S@5$.

6.4 Evaluation of Prescription Analysis

In this subsection, we provide some insights to analyze the guidelines obtained from the cross-culture user study.

6.4.1 On the Cultural Privacy Perception. Due to the limited space, we only displayed the eight categories with the most different and similar privacy perceptions between users from the U.S. and Asia in Tables 4 and 5, respectively. As expected, Table 4 demonstrates the existence of cultural difference regarding privacy perception. Overall, the Americans are more open to share the listed personal aspects, especially their emotions, age and activities at home, as compared to the Asians. As can be seen, American are more open than Asians in terms of revealing their age. Therefore, it is advisable to recommend different actions for users with different cultures, once the privacy leakage is detected.

Table 5 also demonstrates the existence of social consensus between different cultures regarding the privacy perception. Interestingly, we observed that the categories on which users from different cultures achieve consensus are more private, as the

Table 3: Performance comparisons between our TOKEN model and the baselines in S@K and P@K (%).

Methods	S@K			P@K			p-value
	$S@1$	$S@3$	$S@5$	$P@1$	$P@3$	$P@5$	
SVM	2.65 ± 1.09	52.15 ± 4.25	72.01 ± 1.28	2.65 ± 1.09	17.80 ± 2.03	16.53 ± 0.52	2.3e−16
MTL_Lasso	43.99 ± 1.18	73.02 ± 1.30	82.26 ± 0.83	43.99 ± 1.18	27.35 ± 0.56	19.34 ± 0.26	6.9e−7
MTFL	43.75 ± 2.03	73.98 ± 1.03	83.69 ± 0.68	43.75 ± 2.03	27.63 ± 0.51	19.70 ± 0.28	3.1e−3
GO-MTL	43.92 ± 1.29	73.93 ± 1.15	83.45 ± 0.94	43.92 ± 1.29	27.25 ± 0.45	19.40 ± 0.31	2.9e−3
TOKEN	44.37 ± 1.33	74.67 ± 1.38	84.66 ± 0.59	44.37 ± 1.33	28.42 ± 0.57	19.86 ± 0.32	-

Table 4: The eight categories with the most different privacy perceptions between the U.S. and Asia. The percentage of subjects who feel comfortable to share the given personal aspect to each social circle. FA: Family Member; CL: Close Friends; CA: Casual Firends; OU: Outsider Audience.

Categories	the U.S.				Asia			
	FA	CL	CA	OU	FA	CL	CA	OU
emotion: positive emotion	95.0%	97.5%	83.0%	54.0%	75.5%	86.5%	44.5%	21.0%
emotion: negative emotion	88.5%	93.5%	59.5%	36.5%	49.0%	77.5%	31.0%	20.0%
personal attributes: gender	95.5%	96.0%	84.5%	63.5%	75.5%	76.5%	53.0%	32.5%
emotion: general complaints	92.0%	94.0%	83.5%	59.5%	67.0%	79.0%	52.0%	32.0%
personal attributes: age	98.5%	96.0%	74.5%	40.0%	89.5%	79.0%	38.0%	16.0%
activity: activities at home	95.0%	93.0%	61.5%	35.0%	79.0%	68.5%	33.5%	13.0%
neutral statements	98.0%	96.5%	94.0%	85.5%	75.0%	81.0%	70.5%	65.5%

Table 5: The eight categories with the most similar privacy perceptions between the U.S. and Asia. The percentage of subjects who feel comfortable to share the given personal aspect to each social circle.

Categories	the U.S.				Asia			
	FA	CL	CA	OU	FA	CL	CA	OU
healthcare: treatments	96.0%	76.5%	18.5%	5.0%	88.0%	65.5%	14.5%	7.5%
healthcare: health conditions	98.0%	71.0%	17.5%	7.0%	85.0%	65.5%	19.5%	7.5%
life milestones: passing away	95.5%	86.5%	35.0%	12.0%	88.0%	74.5%	31.0%	7.5%
emotion: specific complaints	53.5%	78.0%	28.0%	17.5%	36.5%	68.0%	28.0%	19.0%
location: home address	95.5%	71.0%	5.0%	3.0%	80.5%	73.0%	18.0%	6.0%
location: current location	94.5%	87.5%	31.5%	9.0%	75.0%	77.5%	35.0%	11.5%
personal attributes: contact	95.5%	87.0%	18.5%	3.0%	77.5%	80.5%	27.5%	10.5%
location: places planning to go	95.0%	91.5%	51.0%	21.5%	77.0%	87.0%	39.0%	13.5%

majority of users agreed that these categories should be kept private from the outsider audience, even the casual friends. Besides, from Tables 4 and 5, we observed that, in general, the more sensitive the information is, the closer the social circles that they prefer to access are. However, we also noticed that the disclose norms may not always prefer family members. For example, people prefer to keep their tweets regarding the "negative emotion" and "specific complaints" away from their closest social circle–family members.

6.4.2 On the Gap Between the Intended Audience and Real Audience. First, we introduced two kinds of audience who play key roles regarding information disclosure: the *intended audience* and the *real audience*. The former refers to those audience the post owner has in mind when he/she posts, while the latter corresponds to those who actually have access to the post. In the context of our work, posts in our dataset are actually accessed at least by the outsider audience circle, otherwise we cannot collect them via the Twitter Search Service. Based on our study, we observed that only 10.5% Asian users consider the outsider audience as the intended audience with regard to the *contact* aspect. This verifies the existence of a prominent gap between the *intended audience* and *real audience* of a UGC, which just in turn confirms that it is highly desired to fix the problem of boundary regulation. Otherwise, the user's privacy may be seriously leaked by the gap between the intended audience and the real audience.

7 CONCLUSION AND FUTURE WORK

In this work, we study the problem of boundary regulation by presenting a scheme, consisting of three components: description, prediction and prescription. As to description, we build a comprehensive taxonomy, construct a benchmark dataset, and develop a set of privacy-oriented features. Experiment results shows that LIWC and Sentence2Vector features are the most discriminative features regarding privacy leakage detection. Meanwhile, we found that the privacy leakage via UGCs holds certain temporal patterns. Regarding prediction, we propose a taxonomy-guided multi-task learning model to categorize social posts, which is able to learn both group-sharing and aspect-specific features simultaneously. Experiment results also verify the advantages of taking the proposed taxonomy into consideration in multi-task learning. In terms of prescription, we construct cross-culture guidelines regarding the user's information disclosure norms based on the crowd intelligence via AMT. With these guidelines and the user's background, we can recommend users to accordingly select audience for their posts. Furthermore, we investigate the cultural comparisons pertaining to the user's privacy perception. Overall, we found that the Americans are more open to share their personal aspects, as compared to the Asians. Meanwhile, we also found that the privacy perception of users with different cultures achieves social consensus in terms of certain categories, such as the healthcare-related categories and location-related categories. Besides, we observed the prominent gap between the intended audience and the real audience regarding information disclosure. This in turn verifies the importance of solving the problem of boundary regulation.

Currently, we only explore the simple linear mapping to model the prediction component. However, the complicated prediction mapping may lie in the highly non-linear space. Therefore, we

plan to extend our work towards applying the more advanced neural networks in our context due to their huge success in various domains [22].

ACKNOWLEDGMENTS

This work is supported by the National Basic Research Program of China (973 Program), No.: 2015CB352502; National Natural Science Foundation of China, No.: 61772310, 61702300, and 61702302; the Fundamental Research Funds of Shandong University No.: 2018HW010, and the Project of Thousand Youth Talents 2016.

REFERENCES

[1] Qingyao Ai, Yongfeng Zhang, Keping Bi, Xu Chen, and W. Bruce Croft. 2017. Learning a Hierarchical Embedding Model for Personalized Product Search. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 645–654.

[2] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. 2008. Convex multi-task feature learning. *Machine Learning* 73, 3 (2008), 243–272.

[3] Jing Bai, Ke Zhou, Guirong Xue, Hongyuan Zha, Gordon Sun, Belle Tseng, Zhaohui Zheng, and Yi Chang. 2009. Multi-task learning for learning to rank in web search. In *The 24th ACM International Conference on Information and Knowledge Management*. ACM, 1549–1552.

[4] Amir Beck and Marc Teboulle. 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences* 2, 1 (2009), 183–202.

[5] Joanna Asia Biega, Rishiraj Saha Roy, and Gerhard Weikum. 2017. Privacy through Solidarity: A User-Utility-Preserving Framework to Counter Profiling. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 675–684.

[6] Aylin Caliskan Islam, Jonathan Walsh, and Rachel Greenstadt. 2014. Privacy Detective: Detecting Private Information and Collective Privacy Behavior in a Large Social Network. In *Workshop on Privacy in the Electronic Society*. 35–46.

[7] Rich Caruana. 1997. Multitask learning. *Machine learning* 28, 1 (1997), 41–75.

[8] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *TIST* 2, 3 (2011), 27.

[9] Zhiyong Cheng, Jialie Shen, and Steven C. H. Hoi. 2016. On Effective Personalized Music Retrieval by Exploring Online User Behaviors. In *Proceedings of the International ACM SIGIR conference on Research and Development in Information Retrieval*. 125–134.

[10] Zhiyong Cheng, Jialie Shen, Lei Zhu, Mohan S. Kankanhalli, and Liqiang Nie. 2017. Exploiting Music Play Sequence for Music Recommendation. In *Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI*. 3654–3660.

[11] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning* 20, 3 (1995), 273–297.

[12] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Major life changes and behavioral markers in social media: case of childbirth. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1431–1442.

[13] Valerian J Derlega and Alan L Chaikin. 1977. Privacy and self-disclosure in social relationships. *Journal of Social Issues* 33, 3 (1977), 102–115.

[14] Jianping Fan, Yuli Gao, and Hangzai Luo. 2007. Hierarchical classification for automatic image annotation. In *The International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 111–118.

[15] Jianping Fan, Yuli Gao, and Hangzai Luo. 2007. Hierarchical classification for automatic image annotation. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 111–118.

[16] Hongliang Fei, Ruoyi Jiang, Yuhao Yang, Bo Luo, and Jun Huan. 2011. Content based social behavior prediction: a multi-task learning approach. In *The ACM International Conference on Information and Knowledge Management*. ACM, 995–1000.

[17] Fuli Feng, Liqiang Nie, Xiang Wang, Richang Hong, and Tat-Seng Chua. 2017. Computational social indicators: a case study of chinese university ranking. In *The International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 455–464.

[18] Joseph L Fleiss, Jacob Cohen, and B. S Everitt. 1969. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin* 72, 5 (1969), 323–327.

[19] Yoav Freund, Robert E Schapire, et al. 1996. Experiments with a new boosting algorithm. In *International Conference on Machine Learning*, Vol. 96. ACM, 148–156.

[20] Debasis Ganguly, Dwaipayan Roy, Mandar Mitra, and Gareth JF Jones. 2015. Word Embedding based Generalized Language Model for Information Retrieval. In *The International ACM SIGIR Conference on Research and Development in Information Retrieval*. 795–798.

[21] Shuguang Han, Daqing He, and Zhen Yue. 2014. Benchmarking the Privacy-Preserving People Search. In *The International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM.

[22] Xiangnan He and Tat-Seng Chua. 2017. Neural Factorization Machines for Sparse Predictive Analytics. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 355–364.

[23] Roger A Horn and Charles R Johnson. 1991. Topics in matrix analysis. *Cambridge University Presss, Cambridge* 37 (1991), 39.

[24] Lee Humphreys, Phillipa Gill, and Balachander Krishnamurthy. 2010. How much is too much? Privacy issues on Twitter. In *Conference of International Communication Association, Singapore*.

[25] Lee Humphreys, Phillipa Gill, and Balachander Krishnamurthy. 2014. Twitter: a content analysis of personal information. *Information, Communication & Society* 17, 7 (2014), 843–857.

[26] Melinda L Korzaan and Katherine T Boswell. 2008. The influence of personality traits and information privacy concerns on behavioral intentions. *Journal of Computer Information Systems* 48, 4 (2008), 15–24.

[27] Abhishek Kumar and Hal Daumé III. 2012. Learning Task Grouping and Overlap in Multi-task Learning. In *International Conference on Machine Learning*. 1383–1390.

[28] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[29] Kun Liu and Evimaria Terzi. 2010. A framework for computing the privacy scores of users in online social networks. *ACM Transactions on Knowledge Discovery from Data* 5, 1 (2010), 6.

[30] Huina Mao, Xin Shuai, and Apu Kapadia. 2011. Loose tweets: an analysis of privacy leaks on twitter. In *Workshop on Privacy in the Electronic Society*. ACM, 1–12.

[31] Frank McSherry and Ilya Mironov. 2009. Differentially private recommender systems: building privacy into the net. In *The International ACN SIGKDD Conferences on Knowledge Discovery and Data Mining*. 627–636.

[32] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*. 3111–3119.

[33] Tom M Mitchell. 1997. Machine learning. *Burr Ridge, IL: McGraw Hill* (1997).

[34] Sandra Petronio. 2012. *Boundaries of privacy: Dialectics of disclosure*. Suny Press.

[35] Lee Rainie, Sara Kiesler, Ruogu Kang, Mary Madden, Maeve Duggan, Stephanie Brown, and Laura Dabbish. 2013. Anonymity, privacy, and security online. *Pew Research Center* (2013).

[36] Manya Sleeper, Justin Cranshaw, Patrick Gage Kelley, Blase Ur, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. 2013. I read my Twitter the next morning and was astonished: A conversational perspective on Twitter regrets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3277–3286.

[37] Xuemeng Song, Zhaoyan Ming, Liqiang Nie, Yi-Liang Zhao, and Tat-Seng Chua. 2016. Volunteerism Tendency Prediction via Harvesting Multiple Social Networks. *ACM Transactions on Information System* 34, 2 (2016), 10:1–10:27.

[38] Xuemeng Song, Liqiang Nie, Luming Zhang, Mohammad Akbari, and Tat-Seng Chua. 2015. Multiple social network learning and its application in volunteerism tendency prediction. In *The International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 213–222.

[39] Xuemeng Song, Liqiang Nie, Luming Zhang, Maofu Liu, and Tat-Seng Chua. 2015. Interest inference via structure-constrained multi-source multi-task learning. In *International Joint Conference on Artificial Intelligence*. AAAI Press, 2371–2377.

[40] Yi Song, Daniel Dahlmeier, and Stephane Bressan. 2014. Not So Unique in the Crowd: a Simple and Effective Algorithm for Anonymizing Location Data. In *The International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 19.

[41] Damiano Spina, Julio Gonzalo, and Enrique Amigó. 2014. Learning similarity functions for topic detection in online reputation monitoring. In *The International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 527–536.

[42] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), 267–288.

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

[44] Yulu Wang, Garrick Sherman, Jimmy Lin, and Miles Efron. 2015. Assessor Differences and User Preferences in Tweet Timeline Generation. In *International ACM SIGIR Conference on Research and Development in Information Retrieval*. 615–624.

[45] Simon S Woo and Harsha Manjunatha. 2015. Empirical Data Analysis on User Privacy and Sentiment in Personal Blogs. In *The International ACM SIGIR Conference on Research and Development in Information Retrieval*.

[46] Sicong Zhang, Hui Yang, and Lisa Singh. 2014. Increased Information Leakage from Text. In *The International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 41–42.