On Anonymous Commenting: A Greedy Approach to Balance Utilization and Anonymity for Instagram Users

Arian Askari
Faculty of Computer Science and Engineering
Shahid Beheshti University, G.C.
ar.askari@mail.sbu.ac.ir

Asal Jalilvand
Faculty of Computer Science and Engineering
Shahid Beheshti University, G.C.
a.jalilvand@sbu.ac.ir

Mahmood Neshati
Faculty of Computer Science and Engineering
Shahid Beheshti University, G.C.
m_neshati@sbu.ac.ir

ABSTRACT

In many online services, anonymous commenting is not possible for the users; therefore, the users can not express their critical opinions without disregarding the consequences. As for now, naïve approaches are available for anonymous commenting which cause problems for analytical services on user comments. In this paper, we explore anonymous commenting approaches and their pros and cons. We also propose methods for anonymous commenting where it's possible to protect the user privacy while allowing sentimental analytics for service providers. Our experiments were conducted on a real dataset gathered from Instagram comments which indicate the effectiveness of our proposed methods in privacy protection and sentimental analytics. The proposed methods are independent of a particular website and can be utilized in various domains.

CCS CONCEPTS

• **Security and privacy → Privacy protections**.

KEYWORDS

Anonymous commenting; User privacy; Provider utility; Privacy-utility trade-off; Instagram

ACM Reference Format:
Arian Askari, Asal Jalilvand, and Mahmood Neshati. 2019. On Anonymous Commenting: A Greedy Approach to Balance Utilization and Anonymity for Instagram Users. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3331184.3331364

1 INTRODUCTION

Motivation. With the advent of Web 2.0, one-way communication on internet transformed into two-way communication where users can openly discuss their ideas about news, products, pictures, videos, etc. In most websites where a comment section is available for the users, a user comment is displayed along with the user's real identity which may threaten her privacy. Recent studies indicate why users may prefer to share their ideas anonymously under certain circumstances [4, 9]. When users use their real identity for commenting, posting, etc. they feel a social pressure to conform to the mainstream and fear not being accepted. Fear of sharing personal data may also root from discomfort and harassment caused by other users. Therefore, the possibility to openly discuss one's ideas and opinions via anonymous comments is an important user requirement. Numerous queries regarding commenting without name display have been queried in Google search engine (Fig. 1). Online platforms, such as Instagram and Youtube, publicly utilize user comments for ranking and content recommendation; thus, anonymous commenting may adversely affect their analysis mechanisms in recommendation services. Yet users who are obliged to use their real usernames for commenting may find their privacy threatened, that is to say, there's a conflict between user requirements (privacy protection) and service provider requirements (data analysis). In this paper, we are concerned with two forms of these analytics. First, how many real users have positive or negative opinions on a specific product which is an important analytic for marketers; for instance, *Zendesk*[1] helps the marketers find out whether they have "*a lot of unique commenters, or one or two people commenting a lot*"[2]. Second, how many real users have the same positive or negative opinion on two different products (i.e. correlation); a real-life example is *Brand24* that lets the marketers know how their customers liked their different products, or how they feel about similar products of different companies[3]. Lack of the former deprives service providers of users' real opinion, and lack of the latter makes product recommendation based on user activity difficult.

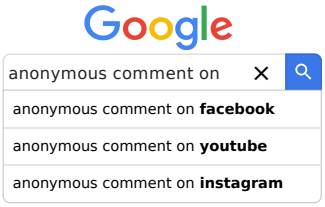

Figure 1: Queries regarding anonymous commenting

User privacy may even be at greater risk due to trading and sharing data by the service providers [2]. Our proposed approach addresses such problems by keeping the identity of anonymous commenters secret even for the service providers.

The research question in this paper is as follows: *How is it possible to receive user comments without their usernames to protect their privacy; considering this anonymity should not adversely interfere with the service providers' analytics*

[1] https://zendesk.com/
[2] http://tiny.cc/zendesk_insta
[3] http://tiny.cc/brand24

SIGIR '19, July 21–25, 2019, Paris, France

ACM ISBN 978-1-4503-6172-9/19/07…$15.00
https://doi.org/10.1145/3331184.3331364

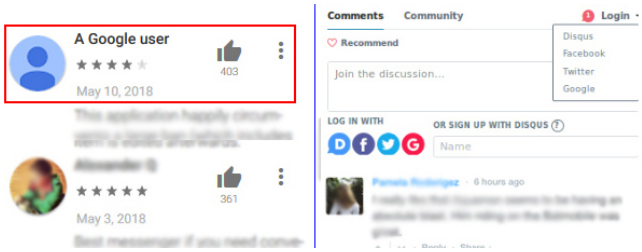

Figure 2: Anonymity in Google Play Store and Disqus

State of the art and its limitations. Currently, anonymous commenting is possible via certain methods. For example, as illustrated on left in Fig. 2, users can comment on a *Google Play Store* product anonymously, going by the name of *A Google User*. This provides an environment where users can openly share their opinion about a product without worrying about their privacy; however, some analytics may be lost.

Disqus[4] is a commenting platform which allows users to comment with or without their real username (Fig. 2, right). If the user wishes to share her real identity, she can integrate her Disqus account with her Facebook, Gmail or Twitter account and comment using her real username. This is useful when the user doesn't mind the privacy matters and is willing to expose her identity. Anonymous commenting is also possible by Disqus in the following ways:

- Commenting without any ties to a particular account using a pseudonym. This approach is similar to *A Google User* and privacy is protected but makes analytics difficult.
- Commenting with a single pseudonym linked to a particular account (i.e. Facebook account), that is to say, a one-to-one mapping of a real user to a pseudonym. Even though this approach makes the mentioned analytics possible, experiments reveal that privacy may be greatly threatened by stylometry and author attribution.

Many studies have been done on user privacy in the past years. For instance, Beigia *et al.* [1] propose a framework for anti-profiling where anonymous searching in search engines is made possible; consequently, analytics for enhancing the service provider ranking is feasible while protecting user privacy. In [2] the authors propose a method for removing profiles disabled by the account owners where statistical analytics by the service provider is not jeopardized, and user privacy of deactivated accounts is also preserved.

Approach and contributions. This paper investigates the problem of anonymous commenting alongside data analytics. Proposed methods must consider two aspects. Firstly, preserving user privacy, and secondly, retaining data analytics. We study the utility loss (how much analytics is deprived) when protecting user privacy by anonymizing its real username.

Remarkable achievements of the present paper are as follows:

- Utilizing mediator accounts for anonymous commenting
- Proposing a numerical measure for assessing privacy threat and data analytic efficiency
- Conducting practical experiments on real data gathered from Instagram

[4]https://disqus.com/

2 CONCEPTS AND NOTATIONS

In anonymous commenting, we deal with three entities:

- The set of items $\{p_1, ..., p_n\}$ which users comment on. Items can be products of a website, Instagram posts, Youtube, etc.
- A set of users $\{u_1, ..., u_k\}$ who comment on items
- A set of comments $\{c_{ij}, ..., c_{np}\}$ where c_{ij} denotes comment by user i on item j

This paper aims to anonymize comment c_{ij} by user u_i in a way that user privacy is preserved and analytics data required by the item owners is accessible.

2.1 Measuring user privacy threat

Suppose user u_i wants to comment on item j with c_{ij}. If this comment is published anonymously, user's privacy is less jeopardized. Previous research [8] show that users follow similar lexical rules and styles when writing texts such as comments. For example, particular words, misspellings and punctuation are powerful signals that can be used to identify the user with high accuracy. User identification is, in fact, a classification problem where the classifier, given a profile as an input, assigns a label to it. The label defines a user identified as the profile owner.

In our problem, the given profile to the classifier can be a comment or a set of comments which we know are published by user u_i. Evidently, the more accurate the classifier is, the higher the chances of exploiting user privacy gets. We use three criteria for privacy assessment: F-measure, recall and precision. SVM classifier with linear kernel is used as a state-of-art method for stylometry identification. We used features proposed in [6] for training the classifier since its dataset is gathered from Facebook which is closer to our work in terms of domain and comment length.

2.2 Measuring utility loss

As mentioned in section 1, by providing anonymous commenting the service providers (SP) may lose some required data for analytics. In this paper, two important analytics conducted by service providers are taken into account, and we aim to assess how much of the required data is lost due to anonymity. Further research can be conducted concerning analytics on location, gender and age of the commenters or temporality of the comments. Two utility loss functions are defined in this paper as detailed below.

Single item. In online services, an important measure (SI) for SPs and third parties who analyze SP data, is what percentage of the audience comment positive (negative) sentiments on a given item. Evidently, comment anonymization must not alter this measure as much as possible. Eq 1 assesses loss quantity in single item mode. SI^+ (SI^-) is the fraction of users commented positively (negatively) on an item, and $SI^{+\prime}$ ($SI^{-\prime}$) is the same fraction after anonymization.

$$UL1 = \frac{\sum_{i=1}^{N}(\frac{SI_i^+ - SI_i^{+\prime}}{SI_i^+})^2 + \sum_{i=1}^{N}(\frac{SI_i^- - SI_i^{-\prime}}{SI_i^-})^2}{2N} \quad (1)$$

where SI^+ and SI^- are calculated using Eq 2 (in this paper, neutral comments are excluded).

$$SI_i^+ = \frac{Pos_i}{T_i},\ SI_i^- = \frac{Neg_i}{T_i} \quad (2)$$

Table 1: Dataset Details

(a) General statistics

Item	Count
Users	581
Posts	435
Comments	86,972

(b) Comment statistics

Item	Count
Avg # of comment/user	140.6
Avg Length	57.9
Avg # of Words	11.5

where Pos_i (Neg_i) denotes the number of users who have submitted positive (negative) comments on post i, and T_i is the total number of commenter for post i.

Pair item. SP needs to know what percentage of users have similar comments (positive/negative) on items i and j simultaneously (PI). This improves accuracy in recommendation strategies for users who have a lot in common. Eq 3 represents PI_{ij}.

$$PI^{+}_{ij} = \frac{Pos_{ij}}{T_{ij}},\ PI^{-}_{ij} = \frac{Neg_{ij}}{T_{ij}} \tag{3}$$

where Pos_{ij} (Neg_{ij}) is the number of users who have submitted positive (negative) comments on posts i and j, and T_{ij} is total number of commenters on both items. When anonymizing the comments, the ideal is that the amount of loss before and after the anonymization would be minimum; thus, utility loss is defined as below:

$$UL2 = \frac{\sum \sum_{i,j,i<j}^{N} (\sum_{i=1}^{N} (\frac{PI^{+}_{ij} - PI^{+\prime}_{ij}}{PI^{+}_{ij}})^2 + \sum_{i=1}^{N} (\frac{PI^{-}_{ij} - PI^{-\prime}_{ij}}{PI^{-}_{ij}})^2)}{2 \times \binom{N}{2}} \tag{4}$$

3 ASSIGNMENT TO ANONYMOUS ACCOUNT

In this section, proposed anonymization approaches are introduced. The ideal approach must preserve user privacy while having the lowest utility loss possible. Approaches are listed as follows:

A Google User (AGU). All anonymous comments are assigned to one account (i.e. *A Google User*, described in section 1).

One-to-One (OTO). In this method, for each user u_i an anonymous account aliased as u'_i is produced, and the commenter uses the alias u'_i for anonymous commenting on an item.

Machine Translation (MT). For each user u_i an anonymous account aliased as u'_i is produced, similar to one-to-one. For higher privacy protection in this method, a round-trip machine translation method (English → French → German →English) [7] is used for anonymizing the comments of users. MT helps to preserve user privacy better than simple one-to-one where methods like stylometry can identify the commenters with high accuracy.

Obfuscation. An anonymous account is built for each user similar to one-to-one, but before publishing the comments, they are anonymized using [5]. To give a basic idea of how obfuscation works, we make a few examples:

- Replacing stop words with alternatives or with a phrase having the same meaning like each → all, some → a few
- Change words from British to American English and vice versa like color ↔ colour, apologise ↔ aplogize etc.

Anonymous account generation. Suppose users $u_1, ..., u_i$ have commented on items $p_1, ..., p_k$. This approach aims to build an anonymous account per real user where the distribution of positive and negative comments on items is similar to its distribution by real users to the extent possible, yet the assigned comments to an anonymous user are written by separate real users. By creating anonymous users similar to the real user, we aim to minimize utility loss and preserve user privacy by combining comments from real users and assigning them to an anonymous account. For this approach, the following two methods are implemented in this paper:

Algorithm 1: Random Anonymous Account Generation

Input: $U = \{u_1, ..., u_k\}, C = \{c_{ij}, ..., c_{np}\}$
Output: Anonymous accounts

$R_u = \emptyset$ ▷ where R_u is the Resolved_Users
$R_c = \emptyset$ ▷ where R_c is the Resolved_Comments
Select a random user u_i from $U - R_u$
repeat
  $C_{u'_i} = \emptyset$
    ▷ where u'_i is the anonymous user equivalent of u_i and $C_{u'_i}$ is the set of comments assigned to u'_i
  $R_u = R_u \cup u_i$
  foreach *comment* $c_{ij} \in C_i$
    ▷ where C_i set of comments of user u_i
  do
    $C' = \{c_{ej} | c_{ej} \in C - R_c,\ sentiment(c_{ij}) == sentiment(c_{ej})\}$
      ▷ where C' is the set of comments on item j with the same sentiment of c_{ij}
    if $C' \neq \emptyset$ **then**
      Select a random comment c_{ej} from C'
      $R_c = R_c \cup c_{ej}, C_{u'_i} = C_{u'_i} \cup c_{ej}$
    else
      $R_c = R_c \cup c_{ij}, C_{u'_i} = C_{u'_i} \cup c_{ij}$
    end
  end
until $U - R_u = \emptyset$

- **Random.** We want to assign comments written by users $u_1, ..., u_i$ to anonymous users. In this method, we select a random user in each iteration and try to build an anonymous user whose comments have the same sense as those belonging to user u_i, but these comments belong to different users. If we fail to substitute user u_i's comment with another one, we assign her own comment to u'_i (algorithm 1)
- **Greedy.** This method resembles the random approach to a great extent, but it has a higher privacy preservation. The difference is that the order of user selection for anonymization is based on the count of comments, and the priority is given to users having more comments. If two users have the same count of comments, one is selected with longer comment length average. The idea behind this approach is that a user with numerous and lengthy comments is more likely to be identified with stylometry methods; therefore, we first anonymize her comments to alleviate the identification risk.

4 EXPERIMENTS AND INSIGHTS

4.1 Dataset

The dataset was created by gathering comments on 435 posts published by Instagram user "arianagrande". English comments with at least 10 characters were selected (via Google Translate API) in data preprocessing. Comments were later filtered by users having at least 90 comments on all the posts in the dataset. Finally, 86,972 comments were gathered. Dataset details are presented in Table 1. Comment sense was identified using the method introduced in [3].

Table 2: Comparison of Anonymization Approaches

Algorithms	Loss1	Loss2	Precision	Recall	F-Measure
AGU	0.064	0.900	17%	5.5%	8.4%
OTO	**0.014**	**0.212**	60.9%	66.9%	63.8%
MT	0.049	0.504	49.3%	50.2%	49.7%
Obfuscate	0.052	0.763	28.6%	33.2%	29.7%
Random	**0.014**	**0.212**	0.6%	1%	**0.7**%
Greedy	**0.014**	**0.212**	0.1%	0.3%	**0.1**%

4.2 Setup

The flow of experiments conducted in the next section for calculating utility loss and privacy threat is as follows[5]:

(1) For each user, half of her comments are anonymized using one of the methods introduced in section 2. There are k real users and m anonymous users in this phase. Each user has some assigned comments detailed as below:
- Set of real users denoted by $K = \{u_1, ..., u_k\}$.
- Set of anonymous users denoted by $A = \{u_1', ..., u_m'\}$.
- Set of real comments $KC_i = \{c_{ij}, ..., c_{ip}\}$ by user u_i where c_{ij} denotes comment by real user u_i on item j.
- Set of anonymous comments $AC_i = \{c_{ij}', ..., c_{ip}'\}$ by anonymous user u_i' where c_{ij}' denotes comment by anonymous user u_i' on item j.

(2) For each set KC and AC, stylometry vectors with features listed in [6] are extracted.

(3) Using SVM classifier, we build a model which estimates the real writer given its AC stylometry vector.

Results and insights. In this section, we compare the performance of proposed methods on our dataset regarding measures related to utility-loss and privacy threat. As mentioned before, utility-loss is calculated based on loss1 (Eq. 1) and loss2 (Eq. 4). Precision, recall and F-measure were used as measures for privacy threat assessment of the algorithms. Evidently, the lower loss1 and loss2 are, the lesser analytics data is lost, and the lower the accuracy of SVM, the higher the user privacy is protected. Table 2 compares our proposed methods based on assessment measures (reports are averaged over 100 runs). We added results of commenting with real username for the sake of comparison in the last row; clearly, utility-loss is ideal while user privacy is threatened to the greatest extent.

A Google User (AGU). This method has the highest loss compared to other methods since all the users' comments are anonymized with the same pseudonym; therefore, SP loses the majority of its analytics. On the upper hand, it preserves user privacy greatly since identification of the real user is practically impossible.

One-to-One (OTO). Compared to *A Google User*, this method has a converse approach. Loss1 and loss2 have greatly decreased (best method for loss alleviation), but the problem is that user privacy is threatened to a great extent, that is to say, identification of the commenter is possible with a 63.8% F-measure. MT-based and obfuscated one-to-one also try to make a trade-off between loss and privacy. Both of these methods improve privacy preservation, but increase loss1 and loss2 compared to one-to-one.

Anonymous account generation. anonymous account generation methods make a more efficient combination of privacy threat and utility-loss compared to stylometry-based methods such as obfuscate and MT which merely improve privacy. Specifically, the greedy method minimized loss1 and loss2 while SVM accuracy for user identification is very low. This methods also outperforms *A Google User* in privacy measure. Comments of different users are displayed under one single account, and extracted features for anonymous comments don't belong to any specific user. This explains why SVM accuracy is lower compared to *A Google User* approach. Greedy method produces the same utility-loss as random approach but lowers privacy threat. The reason is that greedy prioritizes the users having a higher chance of identification with stylometry in the anonymization process. Fig. 3 shows comparison of all the algorithms based on privacy and utility-loss measures.

[5] http://tiny.cc/anonymous_commenting

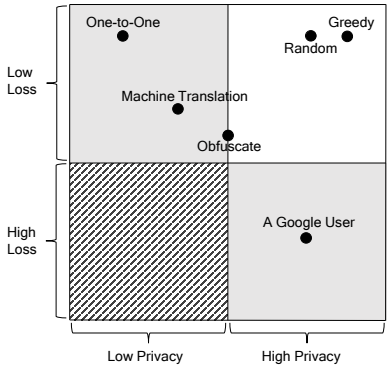

Figure 3: Comparison of Anonymization Approaches

5 CONCLUSION

We introduced the anonymous commenting problem and proposed an alternative approach as opposed to traditional methods such as *A Google User*. In our proposed methods, anonymous comments belonging to real users are assigned to anonymous accounts so that apart from high user privacy preservation, service provider requirements for data analytics are not adversely affected. Our experiments on an Instagram dataset using various algorithms demonstrate that high user privacy protection while retaining precise data analytics is possible. An issue to resolve for future studies is implementing the algorithms in an online environment with parallelization capabilities in order to conduct the algorithms in large scale.

REFERENCES

[1] A. Biega, R. Saha Roy, and G. Weikum. 2017. Privacy through Solidarity: A User-Utility-Preserving Framework to Counter Profiling. In *Proc. of SIGIR '17.*
[2] S. Eslami, A. Biega, R. Roy, and G. Weikum. 2017. Privacy of Hidden Profiles: Utility-Preserving Profile Removal in Online Forums. In *Proc. CIKM '17.*
[3] CJ. Hutto and E. Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proc. of ICWSM '14.*
[4] R. Kang, S. Brown, and S. Kiesler. 2013. Why do people seek anonymity on the internet?: informing policy and design. In *Proc. of SIGCHI '13.*
[5] G. Karadzhov, T. Mihaylova, Y. Kiprov, G. Georgiev, I. Koychev, and P. Nakov. 2017. The Case for Being Average: A Mediocrity Approach to Style Masking and Author Obfuscation. In *Proc. of CLEF '17.*
[6] JS. Li, LC. Chen, JV. Monaco, and P. Singh. 2016. A comparison of classifiers and features for authorship authentication of social networking messages. *CC-PE* (2016).
[7] JR. Rao and P. Rohatgi. 2000. Can Pseudonymity Really Guarantee Privacy?. In *Proc. of SSYM '00.*
[8] A. Rocha, WJ. Scheirer, CW. Forstall, T. Cavalcante, A. Theophilo, B. Shen, ARB. Carvalho, and E. Stamatatos. 2017. Authorship attribution for social media forensics. *IEEE TIFS* (2017).
[9] K. Zhang and RF. Kizilcec. 2014. Anonymity in Social Media: Effects of Content Controversiality and Social Endorsement on Sharing Behavior. In *Proc. of ICWSM '14.*