Privacy-Preserving Text Labelling Through Crowdsourcing

Giannis Haralabopoulos, Mercedes Torres Torres, Ioannis Anagnostopoulos, Derek Mcauley

Privacy-Preserving Text Labelling Through Crowdsourcing

Giannis Haralabopoulos[1][0000−0002−2142−4975], Mercedes Torres Torres[1], Ioannis Anagnostopoulos[2], and Derek McAuley[1]

[1] University of Nottingham, Nottingham, UK name.surname@nottingham.ac.uk
[2] University of Thessaly, Lamia, GR janag@dib.uth.gr

Abstract. The extensive use of online social media has highlighted the importance of privacy in the digital space. As more scientists analyse the data created in these platforms, privacy concerns have extended to data usage within the academia. Although text analysis is a well documented topic in academic literature with a multitude of applications, ensuring privacy of user-generated content has been overlooked. In an effort to reduce the exposure of online users' information, we propose a privacy-preserving text labelling method for varying applications, based in crowdsourcing. We transform text with different levels of privacy and analyse the effectiveness of the transformation with regards to label correlation. To demonstrate the adaptive nature of our approach we also employ a TF/IDF filtering transformation. Our results suggest that total privacy can be implemented in labelling, retaining the annotational diversity and subjectivity of traditional labelling. The privacy-preserving labelling, with the use of NRC lexicon, demonstrates an average 0.11 Mean Spearman's Rho correlation, boosted to 0.124 with TF/IDF filtering.

Keywords: Privacy · Crowdsourcing · Labelling · Natural Language Processing

1 Introduction

Sentiment analysis is a human-centred task, where emotions are uncovered from information. Modern methods can work with almost any type of emotion-evoking information such as multimedia content or images. Modern emotion models rely on simple textual information found in OSNs (Online Social Networks) or online review sources.

Text collections are labelled and analysed to create emotion detection and prediction algorithms [6]. Labelling can happen at paragraph, sentence, or term group level. In lexicon-based supervised learning, words are matched to an emotion using a predefined lexicon (sentiment lexicon) often created through crowdsourcing. Crowdsourcing enables researchers to reach a wide range of non expert individual contributors, using various platforms.

The quality of a lexicon-based learning method depends on multiple factors such as the lexicon, the number of labels, and the model. When dealing with

OSNs, the group of words that requires labelling is an OSN submission as a whole. A commonly used practice, that overlooks user privacy, is when the OSN submissions are provided unaltered to crowdsourcing contributors, which makes the creator of the submission potentially traceable. Data submitted in Social Networks is owned by the Social Network itself and is considered open data for any interested individual. This should not reduce our share of responsibility to ethically handle that data. We consider online data as personal data and therefore our study aims to minimise data exposure.

We propose a method for masking text elements based on specific text properties, emotions in our case. This introduces a layer of privacy between the social media users, whose submissions are used in a crowdsourcing task, and the crowd contributors that annotate these submissions. We assess the feasibility of privacy-preserving labelling based on individual term lexicons. Although lexicon-based methods and individual term labelling are governed by a certain level of decontexualisation and their meaning might be miss-interpreted, they are the simplest ingredient of supervised sentiment analysis.

As mentioned, the transformations we propose are based on textual properties. In sentiment analysis, these properties are the emotions conveyed through text. We demonstrate the effectiveness of masking in emotion labelling, each term corresponds to a range of emotions, and experiment with four different text transformations of varying levels of privacy. We explore the results of these transformations with regards to the emotional diversity contributed and suggest an aggregation of individual term emotions as a validator for sentence labels. The results are compared to usual text annotation, to outline the similarities or differences of subjective privacy-aware labelling versus traditional labelling.

1.1 Motivation and Contributions

As mentioned, Online data is considered open data. If a user publicly submits an item (post, photo, video) to a public domain, the consensus is that everyone can use that data freely. We argue that since users do not explicitly agree for their content distribution, such usage is unethical. Furthermore, as researchers we usually have to expose public data to third parties. This exposure is harmful for both the content creator and the third party. Our study is based on the notion that there should be a layer of privacy in-between OSN users and third parties. For example, Psychology researchers usually have to expose themselves to mentally harmful text data in order to perform a task. Our method can maintain a certain level of information and at the same time provide an alternate visualisation of the data.

We aim to create a method that can produce usable labels for automated tasks. With our data transformation method we protect both the individual content creators and third parties exposed to the content. Our proposed transformation method is an initial implementation of privacy-preserving text analysis. Furthermore, to evaluate sentence labelling 'without prior knowledge or gold standards, we propose a naive aggregation of the per term emotion. Finally, we

demonstrate the flexibility of our transformation method by employing a Term Frequency/Inverse Document Frequency (TF/IDF) term weighing.

2 Related Work

Since our text transformation method is quite novel and no similar study exists, we will focus on the most significant crowdsourcing studies that address any form of privacy concern. OSNs privacy has been extensively studied since the early 2000s, but privacy of OSN users with regards to the analysis of their social media submissions is relatively unexplored. Researchers have assessed user vulnerabilities in social media and their actions and those of their social circle [5], noting that users' privacy may be easily infringed [15] even when scientists and developers use data for fair purposes, such as creating personalized experiences. In [22], the authors propose the segregation of privacy concerns to sets of varying privacy priority.

The privacy paradox, as introduced by Banrnes [1], and its applicability to OSN users is the subject of [3]. The study highlights the correlation of online privacy attitudes with personal privacy attitudes, and concludes that online privacy should no longer be considered as paradoxical to real life privacy. Smart living has brought interconnected devices to our daily lives, along with the need for privacy in the IoT space. At a hardware level, [2] introduces a privacy-enhanced participatory query infrastructure for devices and users.

Privacy-enhancing technologies for analysing personal data are also proposed in [21], which focus on Mobile Crowdsourcing Networks. Spatial crowdsourcing is studied in [14] where authors employ an encryption of coordinates to preserve location privacy in geometry based tasks. Contributors are also engaged to assess privacy, especially in computer vision: in [12] crowd-workers compare blurring, pixelisation, and masking video effects with regards to privacy, and privacy intrusiveness of HDR imaging is studied in [13] with crowdsourced evaluation. Authors of [20] experiment on privacy-preserving action recognition. Crowd sourced OSN data published "as is" poses privacy threats for the participating individuals. The authors of [19] propose a privacy-preserving framework for real-time crowd sourced spatio-temporal data. Databox explores privacy-aware digital life [17] and acts as a personal locally-stored data repository to empower users to manage their personal data.

Privacy and crowdsourcing are the governing themes of our study, while the IoT space is a fitting area of application. With the data-box architecture and data anonymisation in mind [17], we propose a text masking method for the analysis of social media submissions. Our method transforms text to vectors and/or images, challenging the perception of participating workers.

3 Proposed methodology

Our proposed process of text transformation, Figure 1, can be summarised as follows. Given a preprocessed cleaned text corpus, we define the text properties

of interest. We then use a per term annotation of the desired properties to transform each term, and in turn each sentence, to a privacy-preserving format. The transformed sentence is then labelled via crowdsourcing. In cases where per-term annotation labels do not exist, they can easily be crowdsourced as single terms, with no privacy concerns.

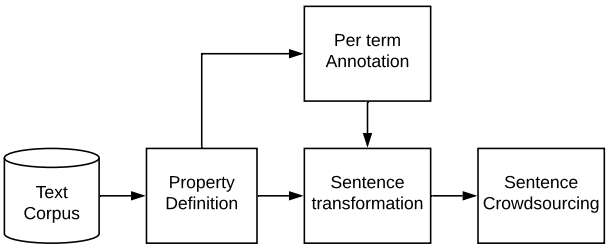


Fig. 1: The process of text transformation

3.1 Lexicon

The sentiment of text can be defined through human labelling. In OSNs, text collections are usually comprised of submissions made from users. The original submission is shown to annotators, who classify the submission according to its sentiment(s). Although public posts in OSNs are considered as public domain, OSN users do not provide an explicit consent for the use of their data in a labelling task, while annotators can easily trace the original author via simple search engine queries. In our study we deal with the transformation of OSN submissions and introduce a privacy layer in-between the crowd and the OSN user.

We propose a privacy-preserving transformation, where words are replaced by their properties. In sentiment analysis applications, the property of interest is emotion. In our study, each word is represented by the emotion it conveys. The Pure Emotion Lexicon (PEL) [11, 7] contains a beyond polarity emotion vector, instead of a single emotion (MPQA, WordNet). The emotional vectors are normalised emotion classification results for each term and correspond to the eight basic emotions, as defined by Plutchik [18].

For instance, the word "normal" received the following annotations in PEL: 0 for anticipation, 0 for sadness, 3 for joy, 0 for disgust, 4 for trust, 0 for anger, 0 for fear and 0 for surprise. Its emotional vector is: [0, 3, 4, 0, 0, 0, 0, 0] and its normalised vector is: [0, 0.75, 1, 0, 0, 0, 0, 0]. We also employ a second lexicon, NRC [16]. NRC is also converted to the same normalised vector format of the eight basic emotions. In total, PEL[3] contains 9736 stems from 17739 terms, while NRC[4] included 3860 stems based on 4463 terms. The emotional distribution within each lexicon can be seen in Figure 2. PEL is dominated by joy annotations, while NRC has a high number of fear annotations.

[3] https://github.com/GiannisHaralabopoulos/Lexicon
[4] http://www.saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

3.2 Privacy

The mathematics formalization of the privacy filter can be described as following. Let d be a text collection with n number of words w.

$$d = [w_1, w_2, ..., w_n] \tag{1}$$

A word w_i is a vector of 8 elements, representing the properties of each word, $e \epsilon [0, 1]$:

$$w_i = [e_1, e_2, ..., e_8] \tag{2}$$

Assuming element e can only have two decimals (i.e. e can have one out of 101 possible values), the number of possible vector permutations for a word w is 101^8.

A document d with n number of words has:

$$(101^8)^n \tag{3}$$

possible permutations. The number of possible three words sentences is more than 10^{48}. In comparison, a 256-bit encryption method has roughly 10^{77} different keys.

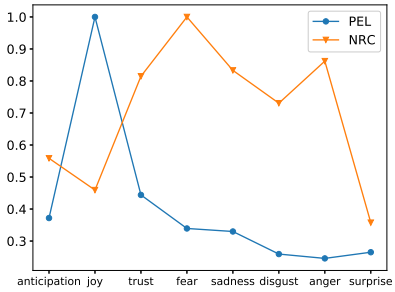


Fig. 2: Distribution of emotions in lexicons

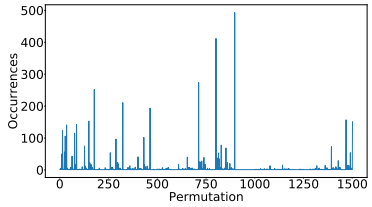


Fig. 3: Distribution of permutations in lexicon

Currently, there are 1502 different emotion vector permutations in the PEL lexicon, distributed as shown in Fig 3.

Given that the emotional vectors are unknown, the permutations for a document d with n number of words from PEL lexicon is:

$$(1502)^n \tag{4}$$

The number of possible three word sentences is almost 10^{10}.

As the number of possible word permutations increases, the identification of the post (and/or the user that submitted it) becomes more and more complex. So that with only three words in a sentence, the number of possible three words sentences is enough to guarantee a high level of privacy. This is without taking into account the image transformation variability, e.g. colour hues, which will be presented in the following subsection.

ant	joy	tru	fear	sad	dis	ang	sur
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0.25	0
0	0	0	0.33	0.33	0.33	0	0
0	0.15	0	0.85	0	0	0	0

Table 1: List of Vectors Transformation

3.3 Transformation

The challenge is to create a representation of text, based on word-property association (emotions in our case) vectors, that will retain labelling performance and annotation diversity. To that end we propose two document transformations that preserve privacy. The proposed text transformations are: List of Vectors (LoV) and Image Vectors (IV).

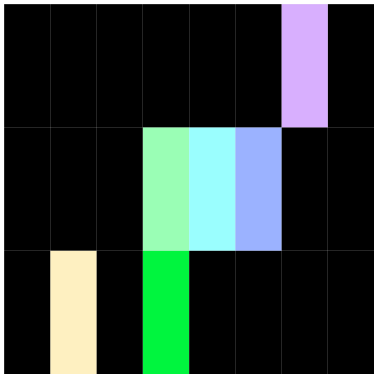


Fig. 4: Image Vector Transformation

Fig. 5: Anger hue range based on vector value

LoV and IoV transformations rely on the vector representation of each word, Equation 2. Let us use the sentence *"They have corruption issues"* to demonstrate the transformation for each method. By using the sentiment vectors we can create an ordered list of representation vectors for LoV, Table 1, where normalized row values correspond to the eight basic emotions (anticipation, joy, trust, fear, sadness, disgust, anger, surprise).

Image Vector Transformation (IV) uses the non zero rows of the LoV transformation to create an image representation of the sentence, Figure 4. Words without at least one emotional annotation, (i.e. "the"), are not drawn. Each vector value is transformed into a certain RGB colour with variable hue. The hue is exponentially analogous to the value it represents, the full hue range for anger can be seen in Figure 5.

The aforementioned example deals with a part of a sentence. Given a document, the analysed parts/sentences can be aggregated, e.g. based on emotional valence, to provide the overall document sentiment. We compare two simple aggregations methods, that can be used to estimate the sentence sentiment based on all the per-term sentiments. These aggregations are: the averaged sum of nor-

malised emotion values per term in a single sentence and the difference of that averaged sum to the mean lexicon emotion.

No Privacy	**Text**
Low Privacy	**Shuffled**
Medium Privacy	**List of Vectors (LoV)**
High Privacy	**Image Vectors (IV)**

Table 2: Privacy Levels

Four crowdsourcing tasks per lexicon are performed. In each task, annotators are presented with only one transformation. E.g. The third task presents LoV tables to annotators, and the annotators must decide on the dominant emotion based on matrices similar to Table 1 with no knowledge of the underlying terms.

4 Experiments

We consider 4 privacy levels in correspondence with the text transformations described above, Table 2. Given an online text submission, the complexity of identifying the user -that submitted the information- increases with each privacy level. We created 4 crowdsourcing tasks per lexicon to analyze the labelling performance of the crowd in diverse privacy settings.

The crowdsourcing tasks were hosted in FigureEight[5] crowdsourcing platform. We selected contributors with higher than B.Sc. education, with native English language skills and the highest level of task completion in the platform. We assess the quality of each participant in our tasks with a subjective quality assurance method that injects objective sentences into the subjective corpus [9]. We also apply a spamming filter at 30% single annotation percentage on all sentence annotations except LoV transformation for "book" and "osn" sources, where 40% and 45% thresholds were applied to retain a sample of statistical significance.

Each sentence in each of the tasks received exactly 10 annotations. Contributors were able to only contribute in one of the tasks, and were excluded from the other three tasks. The use of external crowdsourcing eliminates biases that exist in an internal crowdsourcing task [10]. We ask contributors a simple question "What is the dominant emotion/colour?". Participants in Text had the full text presented to them, while in all other tasks participants could only view the corresponding transformation and not the initial text. In all tasks except IoV transformation task, the available answers are the eight basic emotions as defined by Plutchik [18]. In the IoV transformation task, the available answers are eight colours, based in the circumplex of emotions [18].

Each set consists of 100 sentences with terms contained in both PEL and NRC lexicons. The sentences were obtained from three sources, a book[6] a news

[5] https://www.figure-eight.com/
[6] https://www.gutenberg.org/ebooks/135

site[7], crawled from Reddit[8] and Twitter[9]. These sources will provide diversity in both formality, sentence size and vocabulary.

As the labels and term annotations are provided via crowdsourcing by anonymous contributors and the task is purely subjective, we only assess contributors based on objective annotations of randomly injected terms [9]. The results are analysed in the triad of Distribution, Difference and Dominance.

Each of the four privacy levels requires different thought process during labelling. *No* and *Low Privacy* levels provide contributors with text and no other emotional information, with *Low Privacy* shuffling the words randomly. *Medium privacy* provides numeric values that correspond to the emotional significance of each term, while *High privacy* level only provides a palette of colours where the emotional significance is represented by hue.

As mentioned, emotional labelling is mainly subjective. Thus, we will use *No Privacy* labels as baseline for comparison with the other Privacy levels. However, *No Privacy* labels should not be interpreted as the correct labels nor as the gold standard. We will present two sets of results, based on the lexicons used to create the transformations, PEL or NRC.

4.1 PEL

LoV and IV are created by using lexicon annotation distributions, therefore PEL acts as the transformation agent of a sentence and the resulting annotations follow the PEL distribution, Figure 2. Labels obtained through IV provided an annotation distribution closer to Text than LoV in 19 out of 24 occasions, and in fewer cases outperformed Shuffled.

Simirarly, the annotational difference follows joy annotations of the PEL lexicon, reflected in both LoV and IV transformations. Sentence annotations via Shuffled method present a slight variation in four out of eight emotions. While, sentences from *osn* source have significant positive 'joy' and negative 'trust' differences due to LoV and IV transformations.

The strength of the dominant emotion for each sentence label, as defined by the majority strength of the annotations, is portrayed in Figures 6. The Text and Shuffled transformation have a low dominant emotion agreement, which indicates diversity of opinions, the subjective nature of the annotation task. LoV and IV transformations have high dominant emotional agreement, probably due to differences in the presentation of the task.

Dominant emotion is characterised by low Text and high LoV and IV agreement regardless of source, while Shuffled sentence agreement varies across sources. Sentences from *book* are a good example of that, Fig. 6. Sentences from *news* source have similarly low Text and high LoV agreement, but Shuffled and IV transformations agreements fall within the area of 65% to 90%. Regardless of the source, LoV presents the highest level of agreement followed by IV.

[7] https://open-platform.theguardian.com/
[8] https://www.reddit.com/
[9] https://twitter.com/

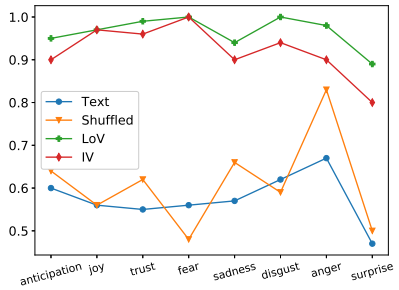


Fig. 6: Dominant emotion agreement for sentences, *book* source

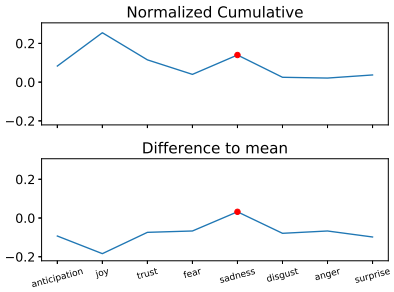


Fig. 7: Per term aggregation based on PEL and Difference to mean PEL aggregation for **sadness** sentences via *IV*

When we aggregate the per term emotion vectors of a sentence, joy is prominent. However, when we calculate the difference of the normalised cumulative sentence emotion to the normalised mean lexicon emotion (diff-aggregation), we can uncover the prominent sentence emotion with high enough certainty, Figure 7. A correlation of the diff-aggregation and the actual sentence annotation is evident in most transformations and sources.

Ant.	Joy	Trust	Fear	Sadness	Disgust	Anger	Surprise
Joy: 0.94	**Joy: 1.89**	Joy: 0.37	Joy: 0.18	Ant: 0.17	Joy: 0.29	Tru: 0.05	**Sur: 0.32**
Ant: 0.62	Ant: 0.58	**Tru: 0.36**	Ant: 0.13	Joy: 0.14	**Dis: 0.12**	Joy: 0.04	Joy: 0.22
Tru: 0.51	Tru: 0.5	Dis: 0.12	**Fear: 0.13**	Tru: 0.11	Tru: 0.12	Ant: 0.03	Tru: 0.19

Table 3: Sentence annotation *(column 1)* and top three IV term aggregation scores for PEL lexicon, *osn* source

Table 3 presents the simple aggregation of terms in relation to the emotion annotation of the sentence for *osn*. The simple aggregation of terms presents high joy annotations, especially in LoV and IV transformations, Table 3. The diff-aggregation corresponds to the Text sentence annotation in 96% of IV transformations. A strong indication of term to sentence relationship, as mentioned in [4].

4.2 NRC

The distribution of sentence annotations for *news* source with NRC transformation is significantly different compared to PEL. 'Joy' was the least annotated emotion in *news* for IV transformation, which is in line with the low number of joy annotations in Text transformation.

The mean annotational difference of LoV and IV to Text is lower than PEL. 'Sadness' in *news* source has the highest difference to the annotations of Text. There is a clear reduction in emotional diversity, with one or -at most- two emotions receiving high number of annotations, while the rest of the emotions are negatively affected. In PEL 'joy' was a key factor in transformation, whereas NRC emotions that positively influence annotations are 'anticipation', 'trust' and 'joy'.

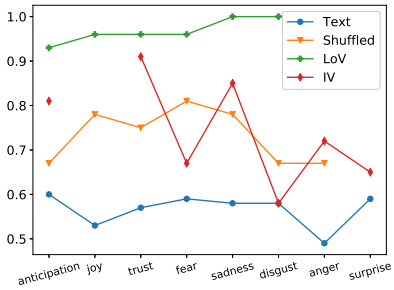


Fig. 8: Dominant emotion agreement for sentences, *news* source

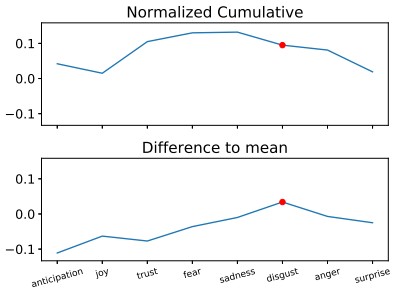


Fig. 9: Per term aggregation based on NRC and Difference to mean NRC aggregation for sentences annotated with **disgust** via *IV*

The annotational agreement for sentences from *book* sources is similar to PEL transformed sentences.The emotion agreement in *news* suggests a greater diversity of opinions, Figure 8.

Ant.	Joy	Trust	Fear	Sadness	Disgust	Anger	Surprise
Ant: 0.52	**Joy: 0.15**	**Tru: 0.6**	Tru: 0.22	**Sad: 0.23**	Tru: 0.11	Fear: 0.14	**Sur: 0.08**
Joy: 0.3	Tru: 0.11	Ang: 0.25	**Fear: 0.18**	Ant: 0.15	Joy: 0.05	Tru: 0.11	Joy: 0.07
Tru: 0.27	Ant: 0.06	Ant: 0.23	Ang: 0.12	Fear: 0.12	Ant: 0.04	**Ang: 0.11**	Tru: 0.07

Table 4: Sentence annotation *(column 1)* and top three IV term aggregation scores for NRC lexicon, *osn* source

Throughout the transformations, the high number of 'trust' annotations in NRC, results in high aggregations for the majority of sentence labels, Table 4. Five out of eight emotions for *osn* source with IV transformation, based on NRC simple emotion aggregation, are highly correlated with the sentence annotation. Diff-aggregation can also be used to determine the most appropriate emotion in more than 93% of the cases, Figures 9.

Source	Book			News			Osn		
Method	Shuffled	LoV	IV	Shuffled	LoV	IV	Shuffled	LoV	IV
PEL	0.2645	0.0678	0.0615	0.1788	-0.0001	0.0069	0.1677	0.0611	0.0567
NRC		0.2065	0.1522		0.0921	0.0988		0.1365	0.0905

Table 5: Mean Spearman's Rho per source and method, compared to Text annotations (Shuffled is independent of lexicon)

4.3 Correlation to traditional labelling

Spearman's Rho correlation of Text annotations against all of our proposed transformations is low when PEL is used as the transformation agent, but greatly improves when we replace PEL with NRC lexicon, Table 5. Shuffled was included as a simple privacy measure, but also demonstrates the highest correlation to Text annotations, despite the fact the Shuffled transformation could exhibit sentiment loss due to rearrangement of terms. The use of a higher quality lexicon

improves the correlation, i.e. improving the transformation agent improves the quality of the labelling process.

5 TF-IDF weighing

ant	joy	tru	fear	sad	dis	ang	sur
0	0	0	0	0	1 (.48)	0	0
0	0	0	0	.06 (.32)	0	0	0
0	.33 (.21)	0	0	0	0	.33 (.21)	0

Table 6: LoV transformation before TF-IDF (after TF-IDF)

To demonstrate the flexibility of our approach, we combine emotion embeddings and a more traditional Term Frequency - Inverse Document Frequency (TF-IDF) calculation. For each term in the corpus we calculate its maximum TF-IDF value, ranging from 0 to 1. This value is then multiplied to the whole emotional vector. For example: the sentence "the insult is not to him but to the law" has an initial LoV transformation as shown in Table 6. After TF-IDF is applied, it has an LoV transformation as shown in parenthesis of Table 6. Similarly, the IV transformation is affected by the TF-IDF weighting. Since each colour is proportionally vibrant to the emotion score, a high TF-IDF enriches colours while a low TF-IDF makes them more subtle. For example "the insult is not to him but to the law" would be transformed as seen in Figure 10. Since none of the terms has a emotion score close to 1, no cell is particularly vibrant in Figure 10b.

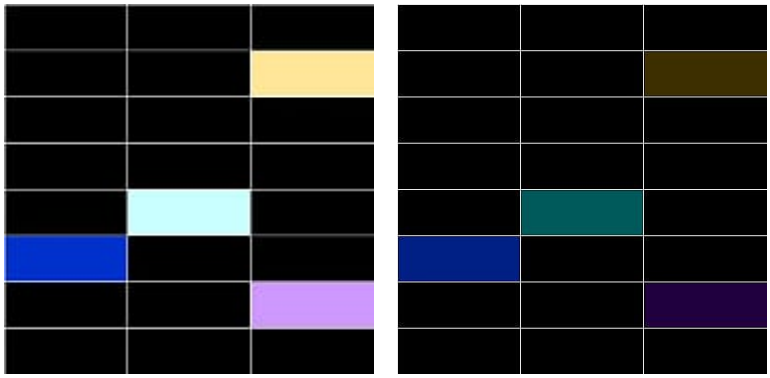
We apply the TF-IDF term weighting in the same set of sentence as before. The exact same tasks are performed in the Amazon Mechanical Turk[10] crowdsourcing platform. Unfortunately, figure eight platform is no longer accessible to researchers. Hence the difference between the shuffled values of Table 5 and Table 7. The TF-IDF approach demonstrates the flexibility of privacy aware crowdsourcing. Furthermore, the mean shuffled correlation is improved by at least 82% (*book* source) and up to 152% (*news* source).

When the PEL lexicon is used as the transformation agent and in *book* and *news* sources, the IV transformation improves the correlation with Text. This indicates that the visual representation could better convey the emotional information when compared to a list of numerical values. Similarly to the previous experiments, the NRC functions as a better transformation agent and all the correlations are better than those of PEL. The most probable explanation for the low correlation values in *osn* is the frequent absence of context from online submissions, which is in turn transferred to the transformations.

5.1 Implications

Based on our results, there exists a positive correlation of the actual labels and the IV/LoV labels. This correlation is lexicon dependent; PEL lexicon is affected by spam, while NRC lexicon miss-represents a range of emotions. Both

[10] https://www.mturk.com/



(a) **(b)**

Fig. 10: IV transformation before(a) and after(b) TF-IDF

the method and the lexicon resource can be further improved. The transformation method is positively affected by a TF/IDF term weighting and correlation is improved by 21.2% on average with NRC and 75% with PEL. TF/IDF is only one type of term weighing that can be applied. Other methods can be used to perform similarity measurements across the dataset or even further enrich the initial term lexicon.

Source	Book			News			Osn		
Method	Shuffled	LoV	IV	Shuffled	LoV	IV	Shuffled	LoV	IV
PEL	0.4815	0.141	0.1535	0.3848	0.0304	0.1005	0.3609	0.0568	0.0147
NRC		0.2407	0.1548		0.1486	0.1311		0.1618	0.0875

Table 7: Mean Spearman's Rho per source and method, compared to Text annotations (Shuffled is independent of lexicon) with TF-IDF

The proposed method is designed to work with any type of information. For example, if researchers are interested in classifying a corpus based on abuse terms, or in a corporate environment where a mental health term lexicon can be used to transform email correspondence to decontextualised and privacy-preserving images. We hope that this study leads to further research in the field of privacy content preservation in crowdsourcing. Up till now, the field is focused in preserving spatial privacy, but our study highlights the feasibility of textual content privacy-preservation as well.

6 Conclusions

We presented a novel approach to privacy-aware labelling that retains subjectivity and is performed through crowdsourcing. The key outcome of our study is: the trade-off between privacy and an as-is presentation is interconnected to the trade-off of agreement and diversity. Text transformations that ensures privacy acts as a curb to contribution diversity, a much needed quality in subjective crowdsourcing tasks[9]. Although manual labelling is not state of the art for

NLP machine learning tasks, labelling of sentences and text is widely used in computer applications [6, 8].

We demonstrated how simple NLP methods, such as TF-IDF weighing, can be used to further improve the correlation results of the LoV and IV transformations. We also presented two naive per term emotion aggregation, capable of acting as a time-effective methods to validate labels. The evaluation of the results is performed via a direct comparison of the privacy aware annotations to non private annotations. We refrain from evaluating the labels in a downstream task, as such an evaluation would add a range of new variables to the experiment.

The lexicons we used contain a low number of emotional permutations and high level of certain emotion annotations, 'joy' for PEL and 'trust' for NRC. Sentences were split based on punctuation, but different splitting methods (e.g. syntactic) should be studied in order to determine the most appropriate approach to privacy-preservation. In addition, the transformation of negation, in a similar text to image scenario, has to be considered. Finally, scaling factor for hues and vibrancy is exponential in our experiment, but different scaling functions can be used to better convey the transformed emotion.

Our proposed text transformation can be applied not only in sentiment analysis tasks. Psychology researchers are often put up against disturbing reports that affect their well-being. A text transformation method can be applied to perform tasks that do not require meticulous study, e.g. a classification of texts based on abuse type. The transition from the traditional text annotation to a more objective visual representation poses challenges to annotators and requesters. Annotators have to adjust their skills to a visual representations, while requesters need to carefully design the transformations in order to preserve the subjectivity of annotations. Emotion transformation is just one of the possible text to property associations that can be used to analyse text. LoV and most importantly IV mask text in a way that provides privacy to the creator and usability to researchers.

References

1. Barnes, S.B.: A privacy paradox: Social networking in the united states. First Monday **11**(9) (2006)
2. De Cristofaro, E., Soriente, C.: Short paper: Pepsi—privacy-enhanced participatory sensing infrastructure. In: Proceedings of the fourth ACM conference on Wireless network security. pp. 23–28. ACM (2011)
3. Dienlin, T., Trepte, S.: Is the privacy paradox a relic of the past? an in-depth analysis of privacy attitudes and privacy behaviors. European Journal of Social Psychology **45**(3), 285–297 (2015)
4. Giatsoglou, M., Vozalis, M.G., Diamantaras, K., Vakali, A., Sarigiannidis, G., Chatzisavvas, K.C.: Sentiment analysis leveraging emotions and word embeddings. Expert Systems with Applications **69**, 214–224 (2017)
5. Gundecha, P., Liu, H.: Mining social media: a brief introduction. In: New Directions in Informatics, Optimization, Logistics, and Production, pp. 1–17. Informs (2012)
6. Haralabopoulos, G., Anagnostopoulos, I., McAuley, D.: Ensemble deep learning for multilabel binary classification of user-generated content. Algorithms **13**(4), 83 (2020)

7. Haralabopoulos, G., Simperl, E.: Crowdsourcing for beyond polarity sentiment analysis a pure emotion lexicon. arXiv preprint arXiv:1710.04203 (2017)
8. Haralabopoulos, G., Torres, M.T., Anagnostopoulos, I., McAuley, D.: Text data augmentations: Permutation, antonyms and negation. Expert Systems with Applications p. 114769 (2021)
9. Haralabopoulos, G., Tsikandilakis, M., Torres Torres, M., McAuley, D.: Objective assessment of subjective tasks in crowdsourcing applications. In: Proceedings of the LREC 2020 Workshop on "Citizen Linguistics in Language Resource Development". pp. 15–25. European Language Resources Association, Marseille, France (May 2020), https://www.aclweb.org/anthology/2020.cllrd-1.3
10. Haralabopoulos, G., Wagner, C., McAuley, D., Anagnostopoulos, I.: Paid crowdsourcing, low income contributors, and subjectivity. In: IFIP International Conference on Artificial Intelligence Applications and Innovations. pp. 225–231. Springer (2019)
11. Haralabopoulos, G., Wagner, C., McAuley, D., Simperl, E.: A multivalued emotion lexicon created and evaluated by the crowd. In: 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS). pp. 355–362. IEEE (2018)
12. Korshunov, P., Cai, S., Ebrahimi, T.: Crowdsourcing approach for evaluation of privacy filters in video surveillance. In: Proceedings of the ACM multimedia 2012 workshop on Crowdsourcing for multimedia. pp. 35–40. ACM (2012)
13. Korshunov, P., Nemoto, H., Skodras, A., Ebrahimi, T.: Crowdsourcing-based evaluation of privacy in hdr images. In: Optics, Photonics, and Digital Technologies for Multimedia Applications III. vol. 9138, p. 913802. International Society for Optics and Photonics (2014)
14. Li, Y., Yi, G., Shin, B.S.: Spatial task management method for location privacy aware crowdsourcing. Cluster Computing pp. 1–7 (2017)
15. Mitrou, L., Kandias, M., Stavrou, V., Gritzalis, D.: Social media profiling: A panopticon or omniopticon tool? In: Proc. of the 6th Conference of the Surveillance Studies Network. Barcelona, Spain (2014)
16. Mohammad, S.M., Turney, P.D.: Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. pp. 26–34. Association for Computational Linguistics (2010)
17. Mortier, R., Zhao, J., Crowcroft, J., Wang, L., Li, Q., Haddadi, H., Amar, Y., Crabtree, A., Colley, J., Lodge, T., et al.: Personal data management with the databox: What's inside the box? In: Proceedings of the 2016 ACM Workshop on Cloud-Assisted Networking. pp. 49–54. ACM (2016)
18. Plutchik, R.: A general psychoevolutionary theory of emotion. Theories of emotion **1**(3-31), 4 (1980)
19. Wang, Q., Zhang, Y., Lu, X., Wang, Z., Qin, Z., Ren, K.: Real-time and spatio-temporal crowd-sourced social network data publishing with differential privacy. IEEE Transactions on Dependable and Secure Computing (2016)
20. Wu, Z., Wang, Z., Wang, Z., Jin, H.: Towards privacy-preserving visual recognition via adversarial training: A pilot study. arXiv preprint arXiv:1807.08379 (2018)
21. Yang, K., Zhang, K., Ren, J., Shen, X.: Security and privacy in mobile crowdsourcing networks: challenges and opportunities. IEEE communications magazine **53**(8), 75–81 (2015)
22. Zheng, X., Luo, G., Cai, Z.: A fair mechanism for private data publication in online social networks. IEEE Transactions on Network Science and Engineering (2018)