Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 1205 - 1211
Received 05 May 2021; Accepted 01 June 2021.

Hybrid Privacy for Social Media Content using Convolutional Neural Networks (CNNs)

Srilakshmi Voddelli[1], Dr. R. Satya Prasad[2],

[1]Research Scholar, Acharya Nagarjuna University,

[2]Professor, Department of Computer Science, Acharya Nagarjuna University.

Abstract: One of the fast-growing fields in computer science engineering is Online Social Networking Sites (OSNS). Every day many people are registering with OSNS to share ideas, make friends, and do other types of activities. With the new users in the OSNS, large data is generated every day by storing users' profiles for further activities. In OSNS, security and privacy are most widely used to prevent attacks on OSNS sever and also personal profiles. It is very important to every user to have privacy for the multimedia content which is accessed by the user. Deep Learning (DL) is the most trending domain nowadays to work on OSNS. In this paper, A Hybrid Privacy for Social Media Content system is introduced to detect the type of image that is uploaded by the OSN user. The proposed system is focused on providing privacy to prevent the user's data from being attacked. The proposed system is integrated with robust pre-processing, Convolutional neural networks (CNNs), and Adaptive Privacy Policy Prediction (A3P). Results show the performance of the proposed system.

Keywords: CNN, A3P, OSNS, DL, Privacy.

Introduction

Online Social Networking Site (OSNS) becomes more popular to communicate the multiple users at the time. This will create more virtual communication among the OSNS users [1]. This social network represents the relationship between various users, companies, and their activities are represented in a social graph. In the graph, all these objects are represented as edges of the graph. By using this platform, the users will create relationships among similar users in terms of views, ideas, and other types of real-life connections among the users [2]. Adaptive Privacy Policy Prediction (A3P) is one of the privacy policy which is used to set the privacy for the user uploaded images. Integration of CNN with A3P gives the accurate and better classification of images and privacy for the users profile data.

OSNS become the most popular and this becomes the culture for a huge number of online users. Merging similar profiles with various communication techniques enables the users to be "in touch" with the OSNs users. Deep learning (DL) is most widely used to identify the various types of images according to the given input. This is mainly focused on classifying the images based on the dataset. The user can upload and share the images and texts in OSNS. Many users check someone's profiles and misuse the profiles by morphing the profile images and photos of OSNS users. DL provides a huge alert on image classification done in OSNS. In this paper, the hybrid privacy algorithm is developed by using the Convolutional neural networks (CNN) and

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 1205 - 1211
Received 05 May 2021; Accepted 01 June 2021.

Adaptive Privacy Policy Prediction (A3P) to predict the accurate image and provide privacy for the input image given by the OSNS user.

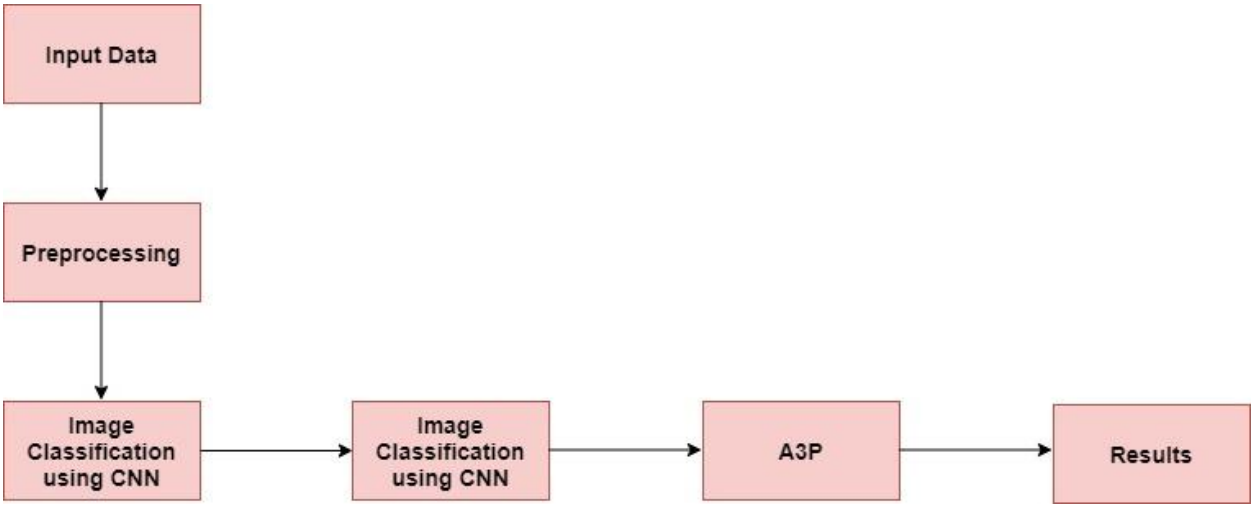

Figure 1: System Architecture

Literature Survey

OSNS is an environment comprising of various substances. These substances incorporate, yet are not restricted to, clients, the OSNS specialist co-op, outsider applications, and sponsors. Notwithstanding, the essential partners of this environment are clients (who get different interpersonal interaction administrations) and OSNS suppliers (who give those long-range informal communication administrations). The protection and security issues bring huge ramifications for clients and OSNS specialist organizations. For clients, potential results mean unseemly sharing of individual data, i.e., spillage, and double-dealing of individual subtleties utilizing dynamic mining, e.g., data linkage [3]. For OSNS administrations, protection and security dangers disturb the appropriate working of the help and harm suppliers' standing.

The scientific classification of protection and security issues in web-based interpersonal organizations is dependent on the partners of the biological system and the tomahawks from which protection and security chances come. As we have as of now referenced, we distinguish two essential partners on internet-based informal organizations: the OSN clients and the OSN itself.

Clients uncover a huge amount of actually recognizable data on OSNS, including physical, mental, social, and special ascribes. For instance, Gross and Acquisti's review [4] show that 90.8% of Facebook profiles have a picture, 87.8% of profiles have posted their introduction to the world date, 39.9% have uncovered telephone number, and 50.8% of profiles show their present home. The concentrate additionally shows that most clients uncover their political perspectives, dating inclinations, current relationship status, and different interests (counting music, books, and motion pictures).

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 1205 - 1211
Received 05 May 2021; Accepted 01 June 2021.

Because of the variety and explicitness of the individual data shared on OSNs, clients put themselves in danger for an assortment of digital and actual assaults. Following, for instance, is a typical danger related to unprotected area data [5]. Segment re-ID was demonstrated to be possible: 87% of the US populace can be remarkably recognized by sex, ZIP code, and full date of birth [6]. In addition, the birth date, old neighborhood, and current home posted on a client's profile are sufficient to gauge the client's government-managed retirement number and in this way open the client to wholesale fraud. Accidental uncovering of individual data brings other web-based dangers, including scratching and collecting [7, 8], social phishing [9], and computerized social designing [10].

Robust Data Preprocessing

This is very important step in the proposed system that helps to process the dataset by removing the blur region, smoothing data and extract the accurate meaningful features from the dataset. In social media images dataset which is collected from kaggle, the pre-processing technique cleans the raw data which is creates the better platform for training models. By using this step, the data is converted to understandable and better format that can be readable by the algorithms.

Some change calculations applying to unique information can be valuable for binning. Normalization strategy is a generally utilized method for a long time learning calculations to determine the issue of various information dispersions. Quantile Transformation (QTF), MinMaxScaler (MMS), and logarithmic calculations scalers are considered to change over information prior to binning. Quantile Transformation is executed to join with EQW in these tests. QTF is considered as a powerful pre-handling procedure since it can decrease the impact of the exceptions in web-based media pictures dataset. Tests in test and approval sets which are more modest or bigger than the fitted reach then, at that point, will be allocated to the limits of the yield dissemination. One more calculation outlined in this review is MinMaxScaler, to make a correlation with QTF and logarithmic calculations. MinMaxScaler changes each element over to a given reach by (1) and (2) equations:

$$X_{std} = \frac{X - X.\min}{X.\max - X_{min}} \quad (1)$$

$$X_{scaled} = X_{std} * (\max - \min) + \min \quad (2)$$

These are the functions do the transformation by using scikit-learn library in python.

Image Similarity Measures

This similarity measure is based on image and Network similarity. User profile is taken as input. User identifying information is extracted from the profile. Similarity index is calculated and if the similarity index is greater than the threshold, then the profile is termed as clone, else normal.

This similarity function is used to measure the two profiles. Given two vectors of attributes, X and Y, the cosine similarity, cos(θ), is represented using a dot product and magnitude as

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 1205 - 1211
Received 05 May 2021; Accepted 01 June 2021.

$$\text{Similarity} = \cos(\theta) = \frac{X * Y}{||X||||Y||} = \frac{\sum_{i=1}^{n} X_i B_i}{\sqrt{\sum_{i=1}^{n} X_i^2}\sqrt{\sum_{i=1}^{n} Y_i^2}}$$

Where, X_i and Y_i are components of vector X and Y respectively.

Dataset and Results

The synthetic dataset is utilized from the various sources. For the training approximately 500 images are taken with different types of categories such as kids, animals, adult, nature and other types of images are present. Over 100 synthetic profiles are taken for the testing purpose over 1500 images is used. The proposed classification is implemented with this datasets. The five parameters sensitivity, specificity, accuracy, recall and F1-Measure are the parameters that can consider overcoming the issues in OSNS. The overall accuracy is high for ensemble classification and advanced privacy for the OSNS.

Algorithm:

Step 1: Initializing the training data (inputs and outputs)-this step consists of 1000 images of different categories for training.

Step 2: Developing and connecting the CNN layers (this consider the preparing of weights, biases, and activation function of each layer). (Internal analysis of every record and attributes and connects the relation between attributes).

Step 3: A loss function is to be developed to assess the prediction error. (if any error occurs such as missing value or unknown value in the record, loss function will work to estimate the predicted error).

Step 4: The training loop is created with network and update its metrics. (Loop is created for updating the data to the network to analyze the metrics).

Step 5: To predict the network accuracy testing data should be access. (The final prediction of data can be done by using the testing data i,e images)

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 1205 - 1211
Received 05 May 2021; Accepted 01 June 2021.

Experimental Results

The experiments are conducted on using python programming language with powerful libraries such as pandas, keras, sklearn and other types of machine learning algorithms are used to get the accurate results. This dataset is huge dataset, so it is very important to maintain system configuration high with 8 GB RAM and 1 TB Hard disk for the maintenance of software's.

Performance Evolution

The performance is calculated by using the measures such as Sensitivity, Specificity and Accuracy, F1-measure. The other count values are such as True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) are used to calculate the above measures.

Precision

This is one of the quality metrics that checks the overall accurate positive predictions done.

$$\text{Presicion} = \frac{\text{No. of TP}}{\text{No. of TP} + \text{No. of TN}}$$

Specificity

This will defined as the equality of the overall negatives that are predicted as negative. This means the original negatives are predicted as positives.

$$\text{Specificity} = \frac{\text{No. of TN}}{\text{No. of TN} + \text{No. of FP}}$$

Accuracy

This will calculate the overall accuracy of the images classified.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Recall

Appropriate when minimizing false negatives is the focus.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 1205 - 1211
Received 05 May 2021; Accepted 01 June 2021.

$$\textbf{Recall} = \frac{\text{TP}}{\text{No. of TP} + \text{No. of FN}}$$

F1 Measure

This will merge the accuracy and recall.

$$\text{F1 Measure} = 2 \times \frac{\text{accuracy} * \text{recall}}{\text{accuracy} + \text{recall}}$$

Table: 1 Classification Results of Hybrid Privacy for Social Media Content

Image types	Precision	Specificity	Accuracy	Recall	F1 Measure
Kids	94.5%	93.1%	98.68%	97.1%	97.31%
Animals	95.1%	94.5%	98.22%	97.34%	98.12%
Explicit	94.5%	96.8%	99.7%	97.87%	97.32%
Scenery	94.5%	95.7%	98.13%	97.23%	97.32%
Adults	94.5%	93.5%	97.9%	97.45%	97.32%

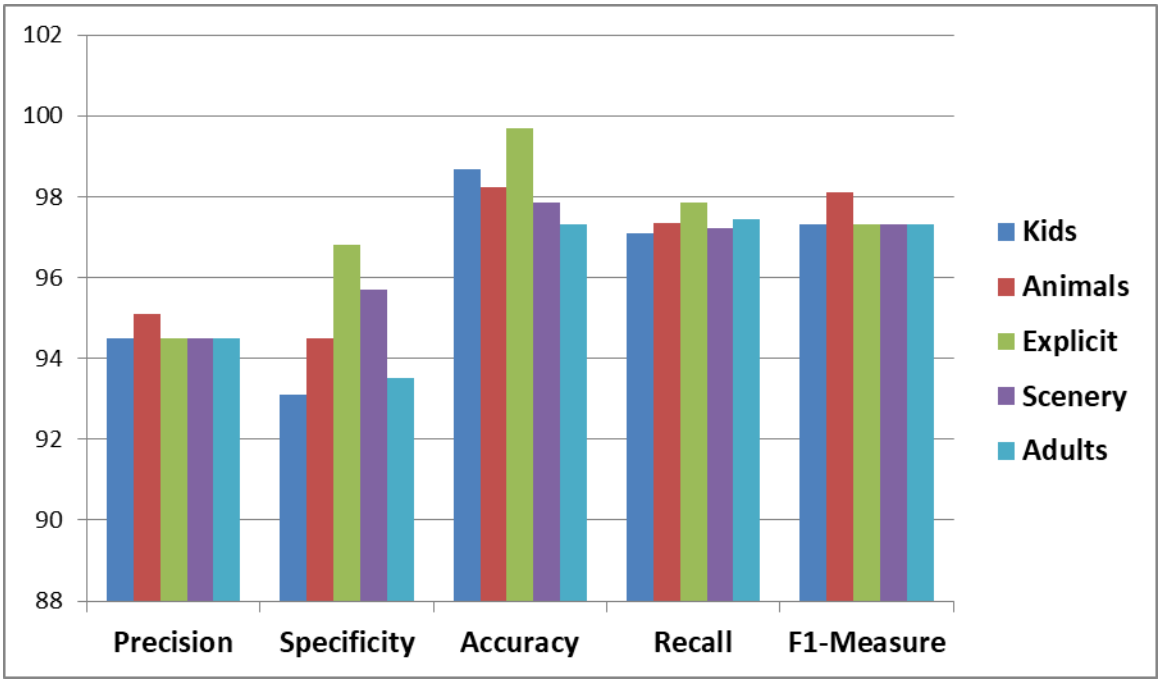

Figure 1: the performance of Hybrid Privacy for Social Media Content

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 1205 - 1211
Received 05 May 2021; Accepted 01 June 2021.

Conclusion

This paper aims to provide privacy for the social media content in the OSNS. This system provides the adaptive classification of images with an accuracy of 98.5%. This is the integration of the CNN and A3P which increases privacy and security with a huge performance. Based on the A3P, the access control is given to the various users to prevent the multimedia content from the attackers. Deep Learning (DL) also plays the significant role in classification of images. In DL, CNN is one of the powerful approache that can identify the images very accurately.

References

1. Boyd, D.M.; Ellison, N.B. Social network sites: Definition, history, and scholarship. J. Comput.-Mediat. Commun. 2007, 13, 210–230.
2. Obar, J.A.; Wildman, S. Social media definition and the governance challenge: An introduction to the special issue. Telecommun. Policy 2015, 39, 745–750.
3. B. Krishnamurthy, Privacy and online social networks: can colorless green ideas sleep furiously? IEEE Secur. Priv. 11 (3) (2013) 14–20.
4. R. Gross, A. Acquisti, Information revelation and privacy in online social networks, in: Proceedings of the ACM Workshop on Privacy in the Electronic Society, ACM, 2005, pp. 71–80.
5. T. Hansen, Social media gives stalkers unprecedented access to victims, 2015,URL: http://www.mcphersonsentinel.com/article/20150112/NEWS/150119927.
6. L. Sweeney, Uniqueness of Simple Demographics in the US population, Carnegie Mellon University, Laboratory for International Data Privacy (2000).
7. J. Lindamood, R. Heatherly, M. Kantarcioglu, B. Thuraisingham, Inferring private information using social network data, in: Proceedings of the Eighteenth International Conference on World Wide Web, ACM, 2009, pp. 1145–1146.
8. T. Strufe, Profile popularity in a business-oriented online social network, in:Proceedings of the Third Workshop on Social Network Systems, ACM, 2010, pp. 2:1–2:6.
9. T.N. Jagatic, N.A. Johnson, M. Jakobsson, F. Menczer, Social phishing, Commun.ACM 50 (10) (2007) 94–100.
10. L. Bilge, T. Strufe, D. Balzarotti, E. Kirda, All your contacts are belong to us: automated identity theft attacks on social networks, in: Proceedings of the Eighteenth International Conference on World Wide Web, ACM, 2009, pp. 551–560.
11. D. Veeraiah and J. N. Rao, "An Efficient Data Duplication System based on Hadoop Distributed File System," 2020 International Conference on Inventive Computation Technologies (ICICT), 2020, pp. 197-200, doi: 10.1109/ICICT48043.2020.9112567.
12. Rao, J. Nageswara, and M. Ramesh. "A Review on Data Mining & Big Data." *Machine Learning Techniques. Int. J. Recent Technol. Eng* 7 (2019): 914-916.
13. Karthik, A., MazherIqbal, J.L. Efficient Speech Enhancement Using Recurrent Convolution Encoder and Decoder. Wireless Pers Commun (2021). https://doi.org/10.1007/s11277-021-08313-6