A Role of Image Processing for Detecting Violations of Facebook Community Standards

[1]A. A. Sattikar, [2]Dr. R. V. Kulkarni
[1]Assistant Professor, [2]Professor & Head
[1]Department of Computer Application,
[1] V. P. Institute of Management Studies & Research, Sangli, Maharashtra, India

***Abstract* - As social networks are growing bigger and more people use them to share more information, it reflects the diversity of the people using it. To balance the needs and interests of a global population, Facebook protects expression that meets the Community standards. There is no way of doing this with 100% certainty. It's difficult but not impossible. Image processing using AI techniques could really helpful in organizing such information and bringing the most relevant pieces to users in a completely authenticated way.**

***Index Terms* – Community Standards, Image Processing, Photoshopped, ANN**

I. INTRODUCTION

Since Facebook gives people around the world the power to publish their own stories, see the world through the eyes of others and connect and share, it has long been a place where people turn to share their experiences and raise awareness about issues important to them. Sometimes, those experiences and issues involve content that is of public interest or concern, such as human rights abuses or acts of terrorism. In many instances, this type of content is bound by Facebook's terms of service and community standards. Facebook community standards govern how the more than a billion Facebook users throughout the world should interact with each other on Facebook. To balance the needs and interests of a global population, Facebook protects expression that does not violate the community standards

II. FACEBOOK SERVICES

Facebook offers a wide variety of products and services, including communications and advertising platforms. Many of these products and services — such as the Facebook mobile app, Messenger, and Paper — are part of Facebook services. Other services, such as Slingshot, Rooms, or the Internet.org app, offer more independent experiences. Worldwide, there are over 1.39 billion monthly active Facebook users (MAUs) which is a 13 percent increase year over year. 890 million people log onto Facebook daily, which represents a 18% increase year over year. Five new profiles are created every second. There are 83 million fake profiles. Every 60 seconds on Facebook: 510 comments are posted, 293,000 statuses are updated, and 136,000 photos are uploaded making it very difficult to manage and monitor the contents.

When users register for Facebook, they must agree to the Web site's terms of service and privacy policies in order to form a profile and use the Web site. Among others, these terms include how and when the Web sites may collect information from a user's profile and computer, how the Web sites track a user's usage, and how they use the information collected from a user's profile. Additionally, these privacy policies describe when other users can view your profile and when and how the Web sites can disclose information to a third party. The privacy policies are mandatory and must be accepted by a user attempting to register for the Web sites.

Hence all of these Services are covered by Data Policy, which describes how Facebook collect, use and disclose user's information. Facebook do it's best to keep it safe, but cannot guarantee it. There is a need of user's help to keep Facebook safe, which includes the following community standards to be followed by users.

III. FACEBOOK COMMUNITY STANDARDS

The Facebook Community Standards page contains the following standards that Facebook user must know:

A. Threats

Any credible threats to harm others will be removed. Facebook remove support for violent organizations.

B. Promoting Self-Harm

Facebook don't allow the promotion of suicide, "cutting," eating disorders, or illegal drug use. Facebook take threats of suicide very seriously and will contact the relevant authorities when they become aware of them.

C. Bullying & Harrasment

Facebook take action when private individuals are bullied or persistently contacted against their wishes. While Facebook encourage meaningful new connections, please keep in mind that contacting strangers or people you've never met in person can be a form of harassment.

D. Hate Speech

Facebook does not tolerate hate speech. It is a serious violation of Facebook terms to single out individuals based on race, ethnicity, national origin, religion, sex, gender, sexual orientation, disability, or disease.

E. Graphic Violence

Inappropriately graphic content will be removed when found on the site. Sadistic displays of violence against people or animals, or depictions of sexual assault, are prohibited.

F. Sex & Nudity

Facebook have a strict "no nudity or pornography" policy. Any content that is inappropriately sexual will be removed.

G. Theft, Vandalism or Fraud

Organizing acts that harm others through theft, vandalism, or fraud is a violation of Facebook terms.

H. Identity & Privacy

Facebook is a community where real people connect and share using their real identities. When you represent yourself accurately on Facebook you are helping to build trust and safety for everyone. Claiming to be someone else, creating multiple accounts, or falsely representing an organization undermines this trust and violates Facebook terms.

I. Intellectual Property

Before sharing content on Facebook, please be sure you have the right to do so. Facebook user only respect for copyrights, trademarks, and other legal rights.

J. Phishing & Spam

Facebook take the safety of FB users seriously and work to prevent attempts to compromise their privacy or security. It also ask that you respect Facebook members by not contacting them for commercial purposes without their consent.

IV. REPORTING TOOL OF FACEBOOK

When people share any content that violates Facebook's terms of community standards and contains violence or threats, bullying or harassment, hate speech, nudity or pornography, it may be removed. In the process of monitoring the accepatability of the content, Facebook takes upon itself the role of a private censor. If user sees something on Facebook that violates our terms, user should report it. User can report posts on their timeline, posts that user was tagged in, or private messages. User can also report a photo, with or without a tag of them. A reporting guide is published to increase the transparency in content review process of Facebook. The guide reveals that Facebook provides for an option where the reportee can appeal to remove content in some cases.

In an effort to quickly and effectively process the millions of reports Facebook receive every day, it has found helpful to contract third parties to provide precursory classification of a small proportion of reported content. In 2012, it was revealed that Facebook outsourced its content reviews to oDesk and provided the reviewers with a 17-page manual which listed what kind of content was appropriate and what was not.

Currently there are dedicated teams throughout Facebook working 24 hours a day, seven days a week to handle the reports made to Facebook. With users all over the world, Facebook handles reports in over 24 languages. Hundreds of Facebook employees are in offices throughout the world to ensure that a team of Facebookers are handling reports at all times.

V. RISK IN REPORTING

In order to effectively review reports, User Operations (UO) is separated into four specific teams that review certain report types - the Safety team, the Hate and Harassment team, the Access team, and the Abusive Content team. When a person reports a piece of content, depending on the reason for their report, it will go to one of these teams. For example, if user is reporting content that contains graphic violence, the Safety Team will review and assess the report.

Facebook is working tirelessly at iterating their reporting system to provide the best possible support for the people who use the site. Due to the complexity of Facebook system and the lack of clarity, the existence of this provision can be misused. As the privacy and security of the people is of paramount importance at Facebook, proper implementation of the policies that Facebook claims to employ, is required along with a systematic method should be used to remove content that is in consonance with community standards.

But while doing so, there is no proper security in the content system and each report can give indication of how much information was on reportee's Facebook page. The moderators can view name of the user who uploaded the reported pages for having abuse content, the subject of the image or person tagged in the photo and also share the information of account holder who did the reporting. Similarly it is tedious job for moderators to manually check these abuse contents, like animal abuse, racism violence and human cut into pieces, which can upset their psychological conditions.

When the growing number of people is established, the platform should provide services that allow the users to find solutions for common problems related with security and privacy. The integration of AI techniques to solve these problems can be a help to achieve intelligent environments offering adaptive behaviors depending on the user's intentions. In order to identify the required steps to contribute in the evolution of these systems, AI techniques can certainly help to improve and design privacy settings to avoid problems like violations of community standards.

VI. RISK IN REPORTING

Image processing is a method to convert an image into digital form and perform some operations on it to extract some useful information or characteristics associated with that image. Usually Image Processing system treats images as two dimensional signals and applies signal processing methods to them. Image Processing forms core research area within engineering and computer science disciplines. Image processing basically includes the following three steps.

- Importing the image with optical scanner or by digital photography.
- Analyzing and manipulating the image which includes data compression, image enhancement and finding patterns from it.
- Preparing output or report that is based on image analysis.

Digital Processing techniques help in manipulation of the digital images by using computers. Any image has to undergo digital technique like Pre- processing, enhancement and display, information extraction as shown in the following figure.

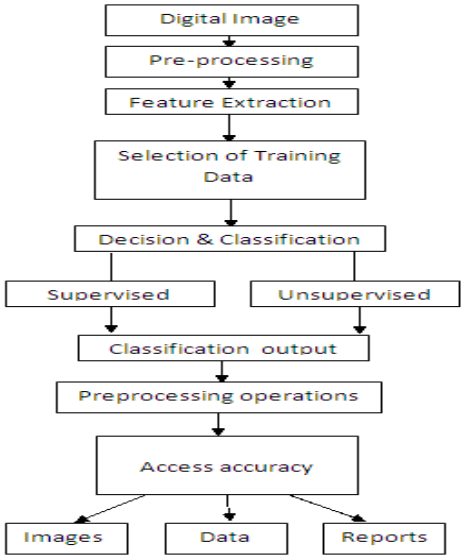

Fig.1 Digital Image Processing Technique

Digital image processing with machine learning techniques allows the use of much more complex algorithms which provides more sophisticated performance at following different tasks.

Classification: Classification is the problem of identifying to which set of categories a new observation belongs on the basis of a training set of data containing observations whose category membership is known.

Feature extraction: When the input data to an algorithm is too large to be processed and it is suspected to be redundant then it can be transformed into a reduced set of features. This process is called feature extraction. The extracted features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data.

Pattern recognition: Pattern recognition algorithms generally aim to provide a reasonable answer for all possible inputs and to perform "most likely" matching of the inputs, taking into account their statistical variation. Pattern recognition systems are in many cases trained from labeled "training" data (supervised learning), but when no labeled data are available other algorithms can be used to discover previously unknown patterns (unsupervised learning).

A self-organizing map (SOM): Like most artificial neural networks, SOMs operate in two modes: training and mapping. "Training" builds the map using input examples (a competitive process, also called vector quantization), while "mapping" automatically classifies a new input vector. A self-organizing map consists of components called nodes or neurons. Associated with each node are weight vectors of the same dimension as the input data vectors, and a position in the map space.

VII. RISK IN REPORTING

Image processing is the method of identifying and detecting an object or a feature in a digital image or video. It is the process of finding instances of real-world objects such as faces, bicycles and buildings in images or videos. Object detection algorithms typically use extracted features and machine learning algorithms to recognize instances of an object category. This concept is used in many applications like systems for factory automation, toll booth monitoring and security surveillance.

In Facebook Community Standard 6.1, it has been mentioned that all photshopped images of people whether they are positive, negative or neutral, are banned. Similarly photos and digital images showing internal organs are also banned. Hence Community standard of Facebook is excellent platform to apply Image processing for detecting violation of these standards.

This article describes how image processing can be applied to detect photoshopped images of people. Photoshopping an image really means combining a lot of layers and redrawing objects to look like real. But it is always tough to combine them together realistically. This bad combination is easy to detect by finding evidences around the edges. Though the lighting discrepancies and fake edges are there, it is very hard to find them with naked eyes. One of the popular methods for this is artificial neural networks. ANN is the term on the method to solve problems by simulating neuron's activities. ANN can be most adequately characterized as

"machine learning model" with particular properties such as ability to adopt and learn, to generalize or to cluster and organize data. ANN has been used in many areas such as interpreting visual scenes, speech recognition, learning robot control strategies, etc.

To identify photshopped images of people, Image processing using neural networks can be implemented in two following phases

A. Detecting Faces

Neural network-based algorithm can be used to detect faces. The algorithm works by applying one or more neural networks directly to portions of the input image and arbitrating their results. Each network is trained to output the presence or absence of a face. To train the neural network a large number of face and non-face images are needed. ANN of two layers with back propagation algorithm can be used for training the system.

B. Detecting Photoshopped Images

To identify photoshopped image, the neural networks can be used by doing pattern recognition, since digital camera's tend to have specific pattern in the picture they create. Most image manipulations will change these patterns which can be recognised. Using neural networks, one can give some indications about what kind of manipulations has taken place. For example, when altering a JPEG image, it is decompressed, loaded into photo editing software, manipulated, then recompressed. Compressing a photo into a JPEG, some information from the EXIF (metadata) data may be excluded. Similarly, when the image is modified, the EXIF data may be completely lost. Using neural networks, one can detect the image that have been manipulated by exhibiting evidence of changes in EXIF information, the secondary compression makes to the image.

VIII. Acknowledgment

The image processing technique based on artificial intelligence like neural network provides sophisticated abilities of intelligent decision making and creative problem solving. Evolutionary computation, clustering, fuzzy sets and pattern recognition are another few examples of artificial intelligent techniques that can be successfully used to solve some relevant problems of detecting violations of community standards of Facebook including photoshopped images. However the benefits of such analysis are even more substantial, including greater confidence in knowledge and ability to predict future outcomes.

References

[1] "10 Facebook Community Standards Facebook Users Must Know" Bottip - SEO Tips, Web Tips and Social Media News Blog

[2] Jessamine Mathew "Content Removal on Facebook — A Case of Privatised Censorship?", Centre for Internet & Society, 16 June, 2014, http://cis-india.org/internet-governance/blog/content-removal-on-facebook

[3] "What Happens After You Click Report", 19 June 2012 https://www.facebook.com/notes/432670926753695

[4] "Privacy & Safety on Facebook , A guide for survivors of abuse" http://nnedv.org/downloads/SafetyNet/NNEDV_FB_Privacy_and_Safety_Guide_2014.pdf

[5] Facebook's Privacy Page: www.facebook.com/privacy

[6] Family Safety Center: www.facebook.com/safety

[7] Basic Privacy Settings & Tools:www.facebook.com/help/privacybasics

[8] Security Tips & Tools: www.facebook.com/help/securitytips

[9] NNEDV's Safety Net Project: www.nnedv.org/safetynet

[10] Safety Net's Technology Safety Materials: www.nnedv.org/safetynetdocs

[11] Jesse D. Kornblum, "Using JPEG quantization tables to identify imagery processed by software", Digital Investigation, Volume 5, Supplement, September 2008, Pages S21–S25

[12] Image Forensics Academy, http://verifeyed.com/documentation/

[13] Facebook Outsourced Abuse Standards/Content Moderation Manual, 24 February 2012

[14] Adrian Chen, "Facebook Releases New Content Guidelines, Now Allows Bodily Fluids , Filed to: SOCIAL NETWORKS , 16 February 2012

[15] A.A. Sattikar, Dr. R.V. Kulkarni (2011), "A Review of Security and Privacy Issues in Social Networking", International Journal of Computer Science and Information Technologies, Vol. 2 (6), 2011, 2784-2787

[16] A.A. Sattikar, Dr. R.V. Kulkarni (2012), "A Role of Artificial Intelligence in Security and Privacy Issues in Social Networking", International Journal of Computer Science and Engineering Technologies, Vol. 2 (1), 2012, 792-795

[17] A.A. Sattikar, Dr. R.V. Kulkarni (2012), "Natural Langauage Processing for Content Analysis in Social Networking", International Journal of Engineering Inventions, Vol. 1 (4), 2012, 06-09

[18] A.A. Sattikar, Dr. R.V. Kulkarni (2013), "A Research on Artificial Intelligence Techniques in Security and Privacy Issues of Social Networking", National conference on Innovative developments and Emerging research Avenues in Software technologies 11 October 2013