RESEARCH ARTICLE OPEN ACCESS

Socially Shared Images with Automated Annotation Process by Using Improved User Preferences

M. Roseline Mercy[1], P. Damodharan[2]
Department of Computer science and engineering, Akshaya College of engineering and technology, India

Abstract
Objectives: The main objective of this research is to increase the semantic concepts prominently as well as reduce the searching time complexity. This is also aimed to ensure the higher privacy with security and develop the accurate privacy policy generation.
Methods: The existing method named as adaptive privacy policy prediction (A3P) is used to discover the best available privacy policy for the user's image being uploaded. The proposed method name as improved semantic annotated markovian semantic Indexing (ISMSI) is used for retrieving the images semantically.
Findings: The proposed method achieves high performance in terms of greater accuracy values.
Application/Improvements: The proposed system is done by using semantic annotated markovian semantic Indexing (ISMSI) approach. ISMSI method is used for identification of similarity as well as semantic annotated images and improves the privacy significantly.
Keyword: Online information services, privacy policy, Meta data, semantic images.

I. INTRODUCTION

Data mining is the process of mining the patterns form data. Generally, data mining is the search for hidden patterns that could be present in huge databases. Data mining scans via a huge volume of data to find out the patterns and correlations between patterns. Data mining approaches are used to extract the image features and policies efficiently. A picture is a collection, or a matrix, of square pixels organized in columns and tuples. The RGB colour model associates especially strongly to the way distinguish colour along with the r, g and b receptors in the retinas. RGB uses additive colour mixing and is the basic colour model used in television or any other medium that projects colour with light. The images are tagged in social network for various purposes and web services are obtained based on user requirements. Web mining is an extensive study area promising to resolve the problems which occur due to the WWW phenomenon. It is an application of data mining methods which is used for mining the knowledge from web data such as documents, hyperlinks among documents and web sites. It is the process of extracting prominent information from the web sites and documents. It consists of text, images, audio, video and records like as lists and tables. The web mining investigation is a joining research area from various examines society, such as databases, information retrieval and artificial intelligence.

Social network is turned into prominent part of daily life as it facilitates to interact along with number of users. Sharing takes place an important role in social media such as facebook, twitter and so on. The users in social networks are sharing images, texts and videos. Many of the content sharing websites permit to enter their privacy preferences. An online social network is permitted the users to make associations and relationships to other users. The association among privacy and a person's social media is complicated and hence the privacy policies are developed to maintain more security for various social media [1].

Alessandro acquisti et.al [2] discussed the topic about online social network growth in this scenario. In modern days the social media is exponentially increased to communicate large volume of information in terms of text, audio and videos among several users. In this scenario, the representative sample of facebook members like social media for colleges and high schools in a foreign education institution and evaluate the survey records to information obtained from the network itself. It is focused on the provision of privacy for the social media users. However it does not suitable for huge dimensional dataset. Shane ahern et.al [3] suggested a methodology for mobile photo sharing by ensuring the privacy concepts efficiently. In this scenario, the analysis focused on the how many flicker (famous web site) users organize their privacy policies for image content. This analysis is consisting of qualitative as well as quantitative analysis. This research scenario is used to prevent the mistakes and increase the awareness of information aggregation. However it has issue with privacy concepts in few cases.

In [4], the research scenario is introduced a recommendation structure to unite image content

along with communities in online social network. This work is a significant to discover the right community for feedback to the end user. This method identifies the images via three kinds of features such as visual features, consumer tag and message history. It provides satisfied precision and recall values. However it provides inaccuracy results in most of the cases. In [5], Ricardo Da Silva Torres discussed to provide image descriptors and relevance feedback. The method introduced is named as content based image retrieval which is used for improving the suitable information systems. The main motivation of this scenario is to maintain image retrieval depends on the content features such as shape, color and texture features. The benefit of this method is probability of an automatic retrieval process which typically requires small amount of time execution for the process and provides meaningful annotation for given database images. However this scenario has issue with information overload.

In [6] the research scenario is introduced models and algorithms for achieving higher privacy on social media network such as face book. The present group of social networks are such as face book and MySpace created a novel class of internet applications named social software. The scenario describes a model for a social network, which is interconnected of entities as well as containers. This is used to promote a community, activity or bigger functions such as network applications and associations. The privacy algorithm is estimating the user's privacy risk caused by privacy settings. In complexity scenario it has issue with providing the higher securities. In [7], Ramprasad Ravichandran et.al discussed about the privacy preferences of social network captures. This research scenario is conducted in the context of location sharing applications, where the consumers are expected to identify constraints under which they are interesting to let other users. The data mining algorithms are such as decision tree and clustering approaches introduced to evaluate the scenario. However it has issue with performance degradation in few cases.

In [8], Katherine starter et.al suggested few strategies along with privacy concepts for this research work. This scenario is focused on the highlights of numerous factors which are restraining the use of privacy systems in social network. It is used to show the users of facebook are attentive that their profiles are public and also used to reveal suitable and safe information. However it has issue with additional serious privacy concerns and designs.

II. Materials and Methods

2.1 FEATURE EXTRACTION

In this section, the features are such as color histogram, texture and SIFT features mined to annotate the images in database. The important features extracted based on the specified images and reduces the unwanted features from the given dataset.

2.1.1 COLOUR HISTOGRAM FEATURE

Color is a significant feature of images. Color features are described topic to a specific color space or model. A numeral of color spaces are used such as RGB, LUV, and HSV. Once the color space is identified, color feature is mined from images or areas. A prominent color features namely color histogram is mined. Color histograms are commonly used to evaluate images. In this gray level difference are used to calculate the histogram of any given image. For this reason the color image is first transformed in to gray level image. Then the histogram values are estimated for gray level differences. According to histogram values, images are mined from the specified database.

In color histogram the number of pixel of specified color is computed the color histogram extraction algorithm entails following three stages.

1. Partition of color space into cells.
2. Association of each cell to a histogram bin.
3. Counting of number of image pixel of every cell and accumulating this count in the viewpoint corresponding histogram bin.

2.1.2 TEXTURE FEATURE EXTRACTION

In this section, texture feature are extracted with the help of SIFT (Scale-invariant feature transform) descriptor. Scale-invariant feature transform (or SIFT) is an algorithm in mainframe idea to identify and illustrate texture features in any given images.

SIFT Descriptors

SIFT based investigation entails discovering salient places in an image and mining descriptors that are unique however invariant to changes in perspective and lighting. The paradigm SIFT interest point detector and the standard SIFT histogram-of-gradients descriptor are utilized. It calculate a single global SIFT descriptor. This universal descriptor is an occurrence count of the quantized narrow descriptors. It employ the clustering approach to cluster a huge group of SIFT descriptors and label each local descriptor along with the identification of the neighbouring cluster centre. The global SIFT descriptor is calculated as

$$SIFT_{GLOBAL} = [t_0, t_1, \dots\dots, t_{k-1}] \quad (1)$$

Where t_i is amount of frequency of the quantized quality features along with label i. SIFTGLOBAL is similar to a term vector in document retrieval. The global SIFT descriptors are normalized to have unit length to account for the varying number of local SIFT descriptors per image.

2.2. Database Updation with HMM annotated images

HMM is used to provide evaluate the parameters of the model from annotated image and caption pairs. Allocating image regions along with caption words in an image and caption pair and calculating the likelihood of a caption word being present in an image.

Let $I = <i_1 \ldots i_t>$ denote segments (regions) in an image, and $C = \{c_1. \ . \ . \ c_N\}$ denote the objects (concepts) present in that image, as specified by the corresponding label (caption) C. For each image region i_t, $t = 1,. \ . \ . \ T$, let $x_t \in IR^d$ represent the color, texture, of the region1. Finally, let V denote the global vocabulary of the caption-words c_n across the entire collection of images.

The scenario is introduced to model the $\{x_t\}$-vectors of an image I as a hidden Markov process, created through an underlying unobserved Markov chain whose states st take values in C. Particularly, each x_t is created according to some possibility density function $f(\cdot|s_t)$ given the state s_t, where s_t itself is a Markov chain along with a known initial state s0 and transition probabilities $p(s_t|s_{t-1})$.

The state sequence is as follows:

$$f(x_1, \ldots., x_T, s_1, \ldots \ldots s_T|s_0)=\prod_{t=1}^{T} f(x_t|s_t)p(s_t|s_{t-1}) \quad (2)$$

Because this stage of aspect is typically not presented in a caption, a hidden Markov model (HMM) is a suitable formalism for calculating the joint likelihood:

$$f(x_1^T,C|s_o)=\sum_{s_1^T \in C^t} \prod_{t=1}^{T} f(x_t|s_t)p(s_t|s_{t-1}) \quad (3)$$

where x_i^T denotes the T-length sequence <x1, .., xT >

2.3. MSI annotated image retrieval

In this module, the user implicitly relates the retrieved images to corresponding query. The Markovian chain evolutions in the charge of the keywords and objective of the scenario are to quantify logical links among keywords. If few user transmits image to his query, where keyword pursues keyword and this occurs m times, then the one step transition possibility is being restructured this procedure builds a Markov chain where each keyword corresponds to a state. Each time keywords emerge in a query, its state counter is advanced; if another keyword follows in the same query, their interstate link counter is also sophisticated. The frequency of the keywords however also the sequencing of these frequencies is both calculated this way.

For the modified MSI approach therefore the steps in constructing the distance table is as follows

1. Parse the text annotation of the images included in the ground-truth, assign an index for each unique keyword and build the one-step transition between keywords probability matrix PG, which is the AMC, considering the annotation of each image as a query related to this image. Similarly to step 1, in order to convert the process to monodesmic, add a small quantity ϵ to all the one diagonal elements (elements lying on the superdiagonal) and subtract it from any random nonzero element in the same line.
2. Perform the eigen decomposition $P_G = VDV^{-1}$ and calculate F_G at the desired n from the following eqn
$$f_i(n) = v_1 + \frac{\tau_i(n)}{n+1} \quad (4)$$
3. Calculate the zero-mean F_G^T by subtracting the mean and perform the eigen decomposition of the covariance matrix of F_G^T as $\sum F_G^T = V_1 D_1 V_1^{-1}$
4. Calculate $B=D_{1k}V_{1k}^{-1}$ where D_{1k} is a square $k \times k$ sub matrix of D1, holding the k largest eigenvalues of $\sum F_G^T$ and V_{1k}^{-1} the sub matrix of the k rows of V_1^{-1}that correspond to these eigenvalues
5. Calculate the reduced k-dimensional $\sum F_G^T$ from $\sum F_G^T = \text{cov}(B^T)$,cov meaning the covariance matrix. At this point we have managed to reduce the dimensionality of $\sum F_G^T$ by projecting on the k principal components. Now we need to project the image vectors in the same space.
6. if A is the matrix having rows the image vectors project the image vectors in the k-dimensional space by $A_k = AV_{1k}D_{1k}^{-1}$ where V_{1k} the sub matrix of the k columns of V1 that correspond to the k largest eigenvalues of $\sum F_G^T$
7. For every pair or rows $r_i r_j$ of A_k calculate their distance by $(r_i - r_j)\sum_k F_G^T (r_i - r_j)$

If we do not need to rise to a power, step 2 above can be omitted. This is the case where the annotation is of unknown origin and there is no Markovian connection to the keyword data.

A3P framework

Users can articulate their privacy preferences about their content discovery preferences with their socially associated users by means of privacy policies. The two key mechanism of A3P-core is such as image categorization and adaptive policy forecast. For every user, their images are primarily categorized depends on content and metadata. Next, privacy policies of each class of images are examined for the policy forecast. Assuming a two-stage scheme is more appropriate for policy suggestion than concerning the common one-stage data mining methods to extract both image features and policies jointly. While a client uploads a new image, the client is waiting for a recommended policy. The two-stage scheme permits the system to utilize the primary step to categorize the new image and discover the candidate sets of images for the succeeding policy recommendation. As for the one-stage extraction approach, it is not able to establish the right class of the new image since its categorization principle

require both image types and policies whereas the policies of the new image are not available yet. Furthermore, merging both image types and policies into a particular classifier would lead to a system which is reliant to the detailed syntax of the policy. Based on the hierarchical extraction concept along with association rule mining we can determine the popular patterns in policies.

6. Improved semantic annotated markovian semantic Indexing (ISMSI)

In this module, annotated images are retrieved semantically based on NLP. Semantic similarity based Image retrieval suggests discovering semantically similar terms using term taxonomies like WordNet.

Improved semantic annotated markovian semantic indexing (ISMSI) procedure

1. The user implicitly transmits the semantically retrieved images to their query.
2. Assume Markovian chain transitions in the order of the keywords the aim of the proposed approach is to quantify logical connections between original keyword as well as semantically matched keywords for the original keyword.
3. Consider the Input image I and Query Q // Query consist of number of keywords K1,K2 and K3 etc
4. Obtain the semantically meaningful keyword for the Keywords for the Original Query by using Natural Language processing tool of Wordnet.
5. Relate Image I to the Query of original keywords with semantically obtained keyword for m times of iterations by using Markov chain transition.
6. Construct and Update transition probability $P_i(k_i, k_j)$ as follows

$$P_i(k_i, k_j) = \frac{N\, Pi\left(ki \cup_{i=1}^{n} Sk_i, kj \cup_{i=1}^{n} Sk_j\right)+m}{N+m} \quad (5)$$

Where $N = M * n(S_k)$

$P_i(k_i, k_j)$ is the transition probability

k_i, k_j denotes keywords in Query

M is the number of original keywords in Query

N is the number of keywords associated with semantically obtained keywords

S_k is the semantically obtained keyword for the original keyword in Query

7. The Aggregate Markovian Chain is constructed of all the queries with its associated keywords asked by all users regardless of the selected images are constructed in this step.
8. Optimization step: The AMC will be used to cluster the semantically obtained keyword space and define explicit relevance links between the semantically obtained keywords by means of this clustering.

III. Results and Discussion

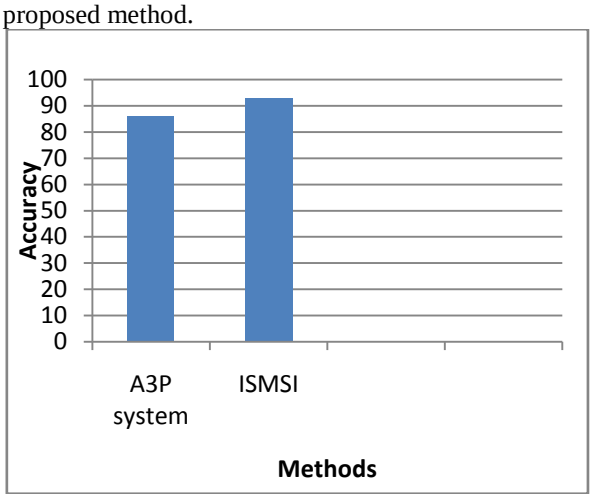
The existing adaptive privacy policy (A3P) technique is used determine the most excellent available privacy policy in the social network while uploading the images. The Improved Semantic annotated Markovian Semantic Indexing method is used to improve the Meta data concept in retrieved images. An experimental result shows that the proposed method achieves high performance in terms of precision, recall, and accuracy.

3.1. Accuracy

Accuracy is defined as the degree of generating the experimental output that is matches with the expected output. The accuracy is calculated by using the following equation

$$\text{Accuracy} = \frac{\text{True Positive} + \text{False Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

In this graph, x axis is taken for two methods of and y axis is taken for accuracy. From the Figure.1 the proposed scenario shows the highest accuracy rather than existing method. ISMSI provides high level semantic annotated Meta data information in proposed method.



IV. Conclusion

In this section, the conclusion decides that the proposed scenario provides superior performance rather than existing scenario. In existing scenario, the method introduced named as adaptive privacy policy prediction (A3P). The main objective of the existing scenario is determining the most excellent available privacy policy in the social network while uploading the images. The A3P system is sued to produce a complete structure to infer a privacy preference depends on the information available for a specified user. But the existing system has issue with inaccuracy results due to the manual creation of Meta data generation. To avoid this issue, in proposed scenario, we introduced the method named as Improved Semantic annotated Markovian Semantic

Indexing (ISMSI) for retrieving the images. It will repeatedly interpret the images using hidden Markov model. Features are such as color and texture are extracted through color histogram and Scale-invariant feature transform (or SIFT) descriptor technique. The parameters of the model are predictable from a group of manually annotated images. From the experimental result, we can conclude that the proposed method of SMSI provides efficient semantic annotations automatically.

V. Acknowledgement

We the authors assure you that, this is our own work and also assure you there is no conflict of interest.

References

[1.] Sangeetha J, A survey on the privacy settings of user data and images on content sharing sites, International Journal of Innovative Research in Computer and Communication Engineering, col.3, Issue 3, 2015.

[2.] Acquisti, Alessandro, and Ralph Gross, Imagined communities: Awareness, information sharing, and privacy on the Facebook, Privacy enhancing technologies, Springer Berlin Heidelberg, 2006.

[3.] Ahern, Shane, et al, Over-exposed?, Privacy patterns and considerations in online and mobile photo sharing, Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, 2007.

[4.] De Choudhury, Munmun, et al, Connecting content to community in social media via image content, user tags and user communication, Multimedia and Expo, 2009,ICME 2009,IEEE International Conference on, IEEE, 2009.

[5.] da Silva Torres, Ricardo, and Alexandre X. Falcao, Content-Based Image Retrieval: Theory and Applications, *RITA* 13.2 (2006): 161-185.

[6.] Maximilien, E. Michael, et al, Privacy-as-a-service: Models, algorithms, and results on the facebook platform, Proceedings of Web. Vol. 2, 2009.