Research Article

Available online at: http://www.ijmtst.com/vol6issue01.html

International Journal for Modern Trends in Science and Technology

ISSN: 2455-3778 :: Volume: 06, Issue No: 01, January 2020

A Novel Approach for Privacy Policy in Inference of User-Uploaded Images Using Contextual Information

T.Suvarna Kumari

Assistant Professor, Department of CSE, Chaitanya Bharathi Institute of Technology, Hyderabad, India

To Cite this Article

T.Suvarna Kumari, "A Novel Approach for Privacy Policy in Inference of User-Uploaded Images Using Contextual Information", *International Journal for Modern Trends in Science and Technology*, Vol. 06, Issue 01, January 2020, pp.-47-52.

Article Info

Received on 03-December-2019, Revised on 26-December-2019, Accepted on 06-January-2020, Published on 20-January-2020.

ABSTRACT

The main objective of this project is to develop methods to predict policy by classifying the images based on content or meta-data in the image context. We implement the proposed protocols and analyze the image and the kind of policy to be mined. We also compare the performance of the proposed meta-data protocols with the content based protocols . Our main contribution is to implement a learning algorithm that is efficient in predicting the appropriate policy for privacy of the uploaded image.

Examine the role of social context, image content, and metadata as possible indicators of users' privacy preferences. Designing a framework which according to the user's available history on the site, determines the best available privacy policy for the user's images being uploaded.

Keywords— *social context, image content,content based protocols*

I. INTRODUCTION

Images are now one of the key enablers of users' connectivity. Sharing takes place both among previously established groups of known people or social circles (e.g., Google+, Flickr or Picasa),and also increasingly with people outside the users social circles, for purposes of social discovery-to help them identify new peers and learn about peers interests and social surroundings.

However, semantically rich images may reveal content-sensitive information. Consider a photo of a student's 2012 graduation ceremony, for example. It could be shared within a Google+ circle or Flickr group, but may unnecessarily expose the students BApos family members and other friends. Sharing images within online content sharing sites, therefore, may quickly lead to unwanted disclosure and privacy violations. Further, the persistent nature of online media makes it possible for other users to collect rich aggregated information about the owner of the published content and the subjects in the published content. The aggregated information can result in unexpected exposure of one's social environment and lead to abuse of one's personal information.

Most content sharing websites allow users to enter their privacy preferences. Unfortunately,recent studies have shown that users struggle to set up and maintain such privacy settings. One of the main reasons provided is that given the amount of shared information this process can be tedious and error-prone. Therefore, many have acknowledged the need of policy recommendation systems which can assist users to easily and properly configure privacy settings.However, existing proposals for automating privacy settings appear to be inadequate to addressthe unique privacy needs of

images , due to the amount of information implicitly carried within

images, and their relationship with the online environment wherein they are exposed. In this project, we propose an Adaptive Privacy Policy Prediction system which aims to provide users a hassle free privacy settings experience by automatically generating personalized policies. The system handles user uploaded images, and factors in the following criteria that influence one's privacy settings of images.

The impact of social environment and personal characteristics: Social context of users, such as their profile information and relationships with others may provide useful information regarding users' privacy preferences. For example, users interested in photography may like to share their photos with other amateur photographers. Users who have several family members among their social contacts may share with them pictures related to family events. However, using common policies across all users or across users with similar traits may be too simplistic and not satisfy individual preferences.

Most content sharing websites allow users to enter their privacy preferences. Unfortunately, recent studies have shown that users struggle to set up and maintain such privacy settings. Therefore, many have acknowledged the need of policy recommendation systems which can assist users to easily and properly configure privacy settings

II. METHODOLOGY

User uploads an image, it is handled as an input query image. The newly uploaded image is then classified using a suitable classification algorithm and the policy as per classification is predicted and displayed to the user. If we use an Evolutionary learning algorithm Social context of users, such as their profile information and relationships with others may provide useful information regarding users' privacy preferences. For example, users interested in photography may like to share their photos with other amateur photographers.

Data miningis a process of extracting useful information from the given data set. Data mining technique includes clustering, classification, regression, association, outlier detection etc. Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both[2].

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyses relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of four types of relationships are sought[4]:

Classes: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.

Clusters : Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities[3].

Associations : Data can be mined to identify associations. The beer-diaper example is an example of associative mining.

Sequential patterns : Data is mined to anticipate behaviour patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes

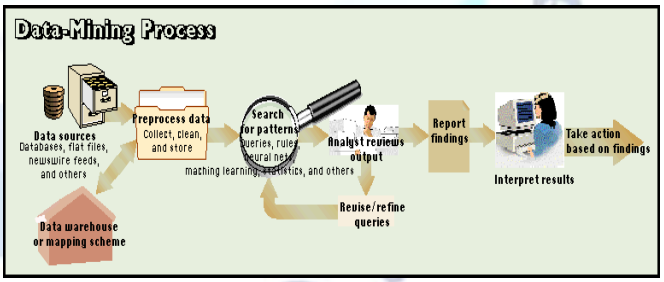

Fig. 1. Data-Mining Process

Data mining consists of five major elements:

- Extract, transform, and load transaction data onto the data warehouse system.
- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in a useful format, such as a graph or table.

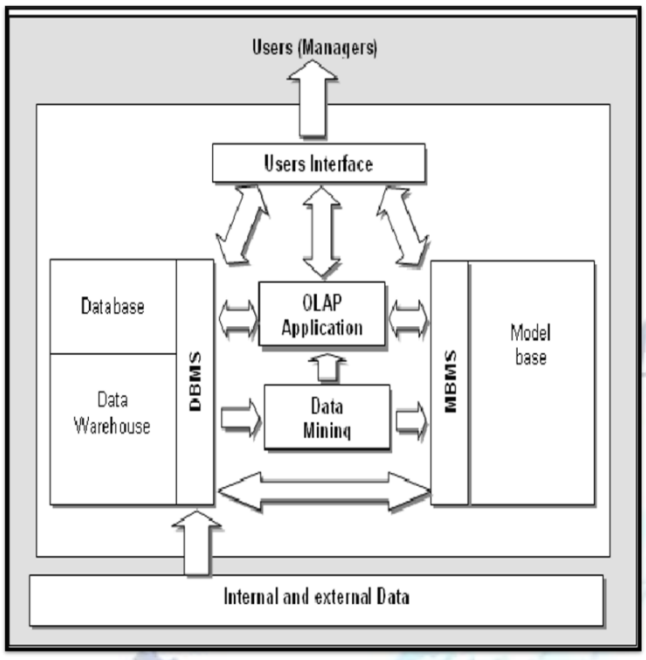

Fig. 2. Architecture of typical data mining system

Different levels of analysis are available:

1. **Artificial neural networks**: Non-linear predictive models that learn through training and resemble biological neural networks in structure.
2. **Genetic algorithms**: Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.
3. **Decision trees**: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection. CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID[2].
4. **Nearest neighbour method**: A technique that classifies each record in a dataset based on a combination of the classes of the *k* record(s) most similar to it in a historical dataset. Sometimes called the *k*-nearest neighbour technique. It supports a huge number of languages and does not narrowed down to just english language which allows a huge weightage over regular porter stemmer. The name stemmer is derived from the developers name Porter as he created a programming language with all new stemming algorithms. Main feature of this Snowball stemmer is that Porter himself stated that Snowball Stemmer is far more efficient than Porter Stemmer which only considers the prefix and sufixes of the words to their root word.
5. **Padding sequences:** Padding in a brief sense means adding extra bits to an original value to normalize it to a specific range. In sentences there might be a wider possibility than few sentences might be short and few might be very humongous . the length of the sentences does not depend on any particular feature. Hence comparing or performing operations, processing these data might in uneven for the model. Hence padding all the sentences or input data into a normalized vector value which can be used as a constant rate for the model and hence can work very efficiently in case of any input. Padding can be done according to our desired length, if less value is used as padded value then the extra data can be truncated and vice versa.
6. **Texts_to_sequence:** Classification of sequences is a key role in predictive modeling problems. These problems mostly revolve around input which have to be united into a specific sequence in order to perform the necessary operations and conclude on one of the outputs. Regularly used to convert a scattered set of data into single sequence so that both of the inputs and outputs are in a singular axis or type and can be compared or processed.The sequence can be formed using NLP modules or any other modules too.
7. **Word Embedding:** Word embedding is a technique where we map each value in the sentence whether it may be a root value or an alternatively used word to that of a real world vector . In other words word embedding is normalization of all the words present in the dataset so as to maintain a range which completely neutralizes the scope for the model and increases the efficiency additionally. The vector values are the high dimensional vector values, the word can map to.
8. **Stopwords:** Terms like "stop words" ,"stop word list" ,"stop list" refer to same thing and is commonly referred as stop words.The term direct to group of words in any language not

only limited to english. Stop words play a very crucial role in certain applications because they add on weight to a sentence by hiding the important words and weighing them down. Because of their wide usage it is applicable to a whole variety of applications. Consider an example , if we perform a search operation of" how to develop a search engine", if the model focuses on "how","to","a" more than "develope","search","engine", it will automatically be over weighed and will tend to give results which may not be related to the search operation . Hence identifying and removal of stopwords add up to a highly efficient solution in terms of a semantic based model.

III. RESULTS

A classification model is tested by applying it to test data with known target values and comparing the predicted values with the known values.

The test data must be compatible with the data used to build the model and must be prepared in the same way that the build data was prepared. Typically the build data and test data come from the same historical data set. A percentage of the records is used to build the model; the remaining records are used to test the model.

Test metrics are used to assess how accurately the model predicts the known values. If the model performs well and meets the business requirements, it can then be applied to new data to predict the future[8].

Accuracy

Accuracy refers to the percentage of correct predictions made by the model when compared with the actual classifications in the test data.

Classification Algorithms

Oracle Data Mining provides the following algorithms for classification

Decision Tree

Decision trees automatically generate rules, which are conditional statements that reveal the logic used to build the tree.

Naive Bayes

Naive Bayes uses Bayes' Theorem, a formula that calculates a probability by counting the frequency of values and combinations of values in the historical data[6].

Generalized Linear Models (GLM)

GLM is a popular statistical technique for linear modelling. Oracle Data Mining implements GLM for binary classification and for regression.

GLM provides extensive coefficient statistics and model statistics, as well as row diagnostics. GLM also supports confidence bounds[9].

CODA (WEB DEVELOPMENT SOFTWARE)

Coda is a commercial and proprietary web developmentapplication for OS X, developed by Panic Sections[10]

The application is divided into six sections (Sites, Edit, Preview, CSS, Terminal, and Books), which are accessed through six tabs at the top of the application. Users can also split the window into multiple sections either vertically or horizontally, to access multiple sections or different files at the same time.

Sites

In Coda, sites are the equivalent of "projects" in many other applications like TextMate. Each site has its own set of files, its own FTP settings, etc. When Coda is closed in the midst of a project and then reopened, the user is presented with exactly what it was like before the application was closed. Another notable feature is the ability to add a Local and Remote version to each site, allowing the user to synchronize the file(s) created, modified or deleted from their local and remote locations[10].

Files

Coda incorporates a slimmed down version of the company's popular FTP client, Transmit, dubbed "Transmit Turbo". The Files portion is a regular FTP, SFTP, FTP+SSL, and WebDAV client, where the user can edit, delete, create, and rename files and folders.

Editor

The editor in Coda incorporates a licensed version of the SubEthaEdit engine, rather than having a custom one, to allow for sharing of documents over the Bonjour network. Coda also has a new Find/Replace mechanism, which allows users to do complex replaces using a method similar to regular expressions.

Coda also recognises specially-formatted comment tags in many syntaxes, called *bookmarks*, which appear in a separate pane beside the editor called the Code Navigator. Bookmarks allow the user to jump to the corresponding line of text from

anywhere in the editor by clicking on the link in the Code Navigator.

Plug-ins

Coda 1.6 and later supports plug-ins, which are scripts usually written in command line programming languages like Cocoa, AppleScript, Perl, or even shell scripting languages like bash, that appear in Coda's menu bar and do specific tasks like appending URLs or inserting text at a certain point. Plug-ins can either be written using Xcode or through Panic's free program, the Coda Plug-in Creator.

Command-line utility

Coda does not come with its own command-line utility. Instead, a third-party utility such as coda-cli can be used.

The IDE CODA was used to develop the working project and deployed on to windows operating system.

Fig.3. Home Screen

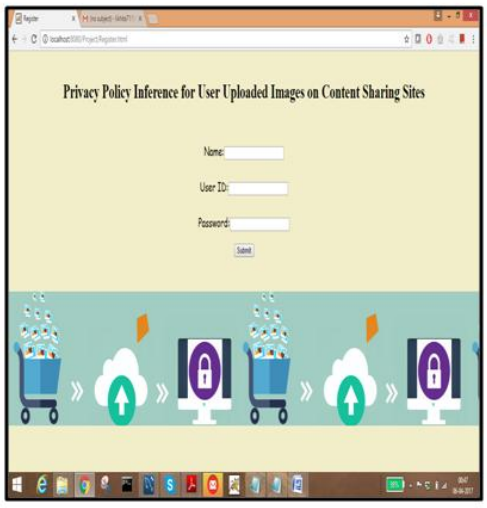

Fig 4. Registration Page

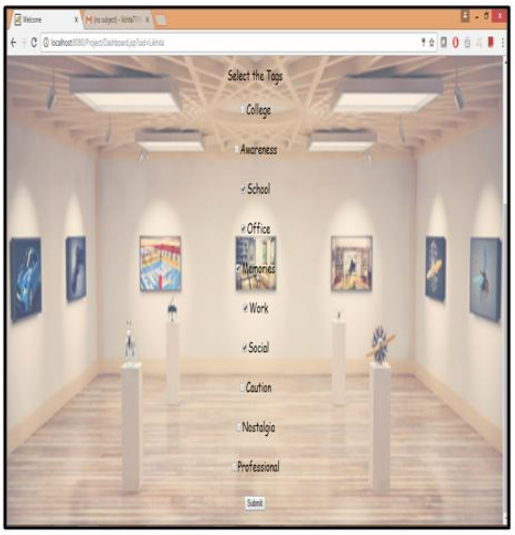

Fig 5. User uploads the image and tags

Fig.6. Depicting the predicted policy

IV. CONCLUSION AND FUTURE SCOPE

We have proposed an Adaptive Privacy Policy Prediction system that helps users automate the privacy policy settings for their uploaded images. The system provides a comprehensive framework to infer privacy preferences based on the information available for a given user. We also effectively tackled the issue of cold-start, leveraging social context information.

Our solution relies on a meta-data classification framework for image categories which may be associated with similar policies, and on a policy prediction algorithm to automatically generate a policy for each newly uploaded image, also according to users' social features.

The limitation of this project is that it requires an initial set of user's policy settings to predict the privacy policy with a considerable accuracy. The

project could further be extended to predicting the policy by processing the image and studying the image features along with the meta-data of the image.

REFERENCES

[1] https://ieeeprojectscbe.wordpress.com/2015/10/07/privacy-policy-inference-of-user-uploaded-images-on-content-sharing-sites-2/

[2] http://www.dataminingblog.com/top-five-articles-in-data-mining/Isabelle Guyon and André Elisseeff

[3] http://www.dataminingblog.com/top-five-articles-in-data-mining/A.K. Jain, M.N. Murty and P.J. Flynn

[4] http://www.dataminingblog.com/top-five-articles-in-data-mining/Usama Fayyad, Gregory Piatetsky-Shapiro and Padhraic Smyth

[5] http://www.csre.iitb.ac.in/~avikb/GNR401/DIP/DIP_401_lecture_7.pdf

[6] A. Acquisti and R. Gross, "Imagined communities: Awareness,information sharing, and privacy on the facebook," in Proc.6th Int. Conf. Privacy Enhancing Technol. Workshop, 2006,pp. 36–58.

[7] R. Agrawal and R. Srikant,"Fast algorithms for mining association rules in large databases," in Proc. 20th Int. Conf. Very Large Data Bases, 1994, pp. 487–499.

[8] S. Ahern, D. Eckles, N. S. Good, S. King, M. Naaman, and R. Nair,"Over-exposed?: Privacy patterns and considerations in online and mobile photo sharing," in Proc. Conf. Human Factors Comput.Syst., 2007, pp. 357–366.

[9] M. Ames and M. Naaman, "Why we tag: Motivations for annotationin mobile and online media," in Proc. Conf. Human Factors Comput. Syst., 2007, pp. 971–980.

[10] https://en.wikipedia.org/wiki/Coda_(web_development_software)

[11] J. Bonneau, J. Anderson, and G. Danezis, "Prying data out of a social network," in Proc. Int. Conf. Adv. Soc. Netw. Anal. Mining., 2009, pp.249–254.

[12] https://www.javacodegeeks.com/2014/05/10-articles-every-programmer-must-read.html.