MINI PROJECT REPORT

UNINTENTIONAL PERSONAL INFORMATION LEAKAGE ON ONLINE SOCIAL NETWORKS

MEMBERS :

Basireddy Durga Shruti (IIT2016019)

Garima Chadha (IIT2016020)

Vaibhav Srivastava (IIT2016034)

Niharika Shrivastava (IIT2016501)

MENTOR:

Dr. Bibhas Ghoshal

DATE:

11/09/2018

Introduction

Due to privacy issues of user information on the internet, and various threats involved in exploiting user data for applications in the field of advertising, business intelligence, and forensics, netizens have restricted providing their personal details to a bare minimum.

But, predicting and/or determining certain social attributes of users on the internet has many positive impacts. Advertisements use the concept of data retargeting and remarketing to provide users relevant recommendations based on their interests and choices. Enhancement of search engine results for queries like “neighboring eateries” or “fun events” is made possible because of predicting trivial user location or age.

In this project, we devise methods to predict various attributes that have not been publicly mentioned in user profiles such as gender, interests/hobbies, and occupation using Twitter as our online social network (OSN). This will in turn help us to establish various use cases for the new user profile generated.

Literature Review

A well known example is Facebook, which reached an achievement with 1 billion users in a single day on August 24, 2015 [1]. The Like and Share buttons are viewed over 22 billion times daily across more than 7.5 million websites [2]. A recent study revealed that 30% of Americans get their news on Facebook [3]. Facebook was criticized for being indifferent to user privacy in the past decade [4].

When it comes to OSNs, information exposure is usually a plus or even a must for users to join a new community [5]. An OSN usually encourages users to expose personal information because self-information disclosure can build trust, strengthen the ties between people, and bind romantic relationships or friendships [6,7].

The usage of OSNs usually raises serious concern about the overall privacy, and it is essential to protect personal identifiable information (PII) [8], which alone or combined with other public information, can be used to distinguish or trace an individual's identity [9].

Motivation

With the growing use of internet and in turn, social networking sites, communication flow between users has increased extensively in the form of blogs, wall posts, messages, connections, broadcasts, etc. Naive OSN users are unaware of the hidden patterns in their public information that might reveal unintentional or involuntary data.

In this project, we devise a method to reveal certain user attributes such as gender, user interests, and occupation. This data can be utilized in various fields of interests like improving search results, marketing purposes, and sentiment analysis. It also points out the extent of vulnerability a user is exposed to unintentionally, which can help them stay clear of cyber crime.

Problem Definition

To devise a method to find unknown user attributes using their publicly available data on an OSN (Twitter) which can be further used for enhanced user experience on the web.

Given: $x \rightarrow length\ of\ set\ of\ public\ user\ attributes\ of\ a\ particular\ OSN$

$n \rightarrow length\ of\ set\ of\ total\ attributes\ that\ can\ be\ found$

$k = n - x$

$k \rightarrow length\ of\ set\ of\ unknown\ attributes$

To find: Set of k unknown attributes of the user using x known attributes.

Technologies to be used

- Python modules (NLTK, pandas)
- Data Mining
- Twitter API (Tweepy)
- Hadoop, Flume
- NLP

Methodology

This section comprises of the steps to be taken for the implementation of this project. A diagrammatic layout of the process has been depicted in the flowchart shown below (Fig. 1).

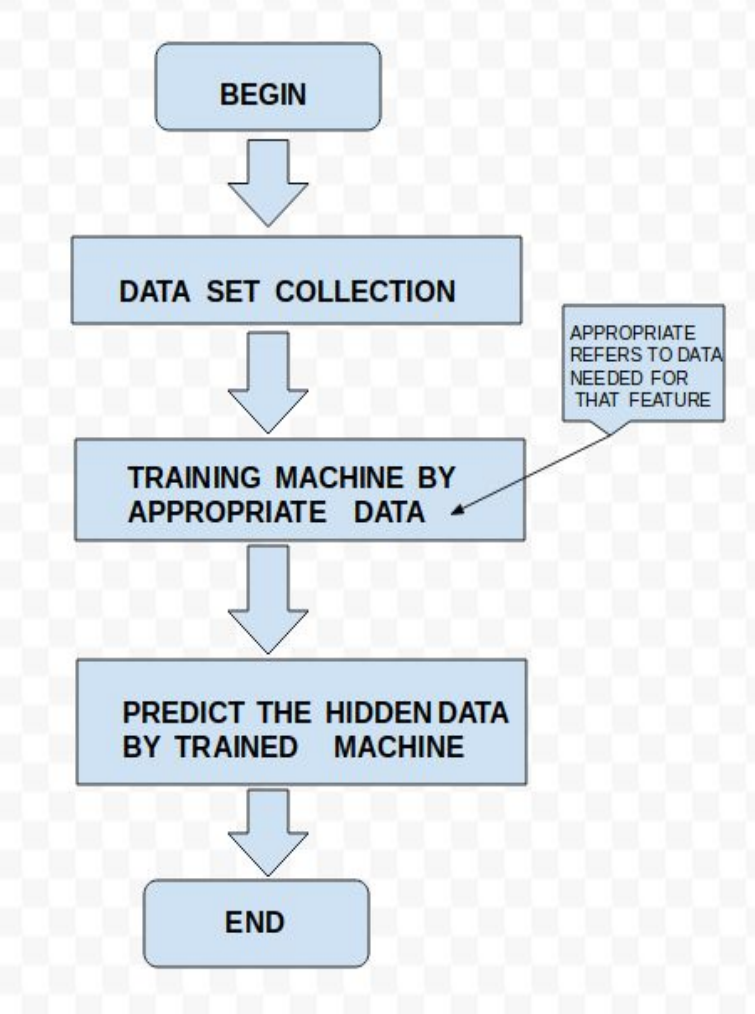

Fig. 1

- **RETRIEVAL OF DATA** - Dataset is collected by making use of the Twitter API (Tweepy). As a part of this process, a Twitter application is created with the corresponding API Key, API Secret Access Token, Secret Access Token. This dataset is collected on a Hadoop Server. Further, queries can be made on this data according to requirement.

- **<u>PREDICTING USER ATTRIBUTES IN TWITTER</u>** - The Twitter dataset is explored to predict the gender, interests and occupation of the user. The contents of the tweets, name, and bio of a user have been analysed to predict these attributes.

  The process can be depicted by the following diagram (Fig. 2):

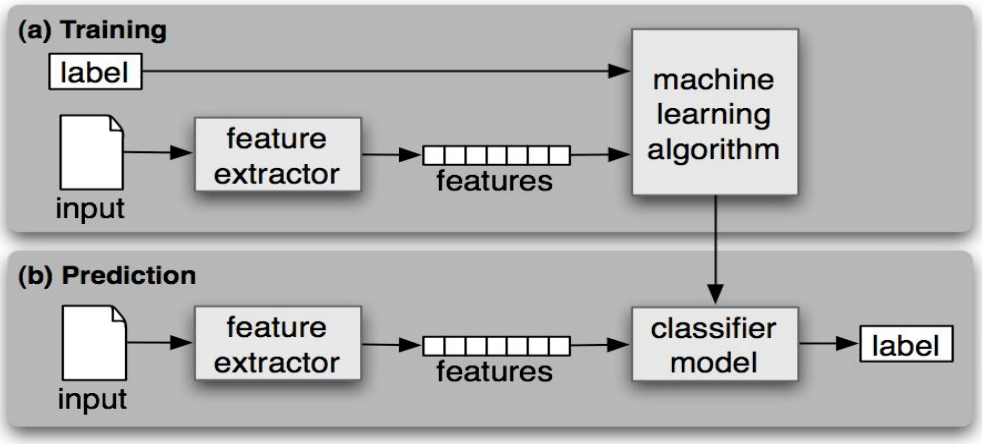

Fig. 2

A) GENDER -

a) As an initial approach, the features of the name of a user are studied to find some pattern. The observations made reveal that **female names** have a greater tendency to end with vowels whereas **male names** generally end with consonants. Along with this, the last letter, the last two letters and the last three letters of the name are stored as part of the **feature vector** of a particular user. These features are used to train the classifier and are considered to be independent of each other. External data is used for training the classifier. **NAIVE BAYES CLASSIFIER** has been used to train and test the data. This classifier uses the Naive Bayes Theorem to find the probability of a user being male or female. We use the inbuilt ***nltk* library** in python to use the Naive Bayes Classifier. The **Bayes Theorem** can be stated as follows :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

A - MALE/FEMALE
B - ELEMENTS OF FEATURE SET.

b) A limitation of above approach is that few users don't reveal their actual names on social media. As a result, accurate gender prediction is not possible. Therefore, we shift to another approach that involves using the tweets of a user to predict gender. In this model, a combination of **POS N-Grams** and **TF/IDF** (Term Frequency - Inverse Document Frequency) has been used.

c) As a third approach, we shall use the bio of a twitter user to look for keywords such as (he/him), (her/she), etc. to determine their genders.

An example of type (a) is shown below(Fig. 3).

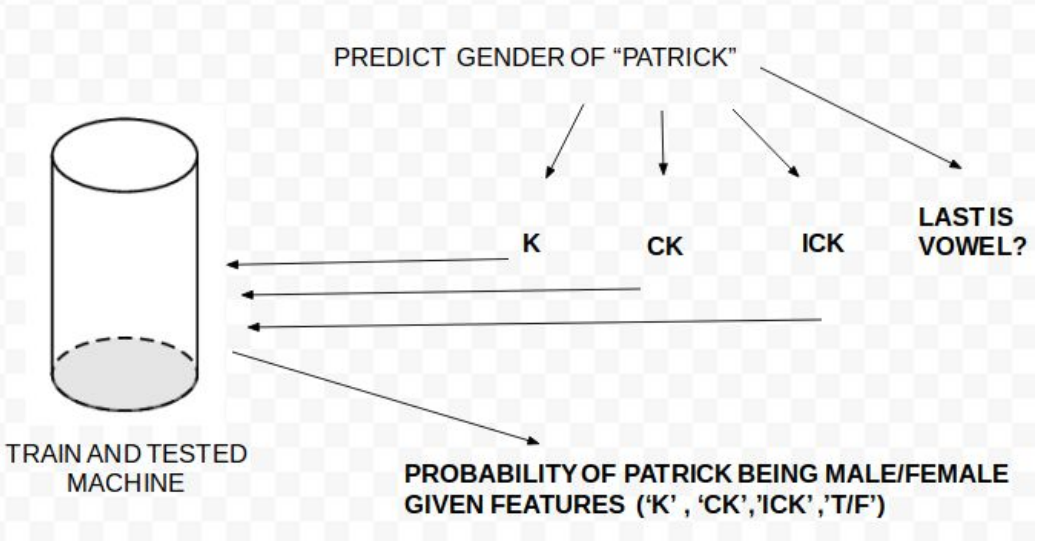

Fig. 3

B) INTERESTS - A set of interests like music, dance, sports, reading, politics etc. is created. For each label belonging to this set, a bag of words having **unigram, digram and trigram phrases** related to concerned label is created. The tweets are then scanned using **TF/IDF** to find the most significant words. This approach may be further extended to predict the interests.

C) OCCUPATION - A set of occupations like artist, health, corporate etc. is created. For each label in this set, a bag of words having n gram phrases related to the label is created. Again, the **TF/IDF** values are used to find the most important words. An extension of this approach may be made to predict the occupation of a user.

Conclusion

- In this project, we have explored the task of predicting user attributes in social media. These attributes can have applications in various fields.
- This project can further be extended to incorporate the following functionalities:
    A) The result of a particular query on search engines can be modified according to user interests.
    B) It is easy to place advertisements if the interests of a user are known beforehand.
    C) Using the features of a particular user, we can establish a mapping between two distinct OSN's, i.e., a user in one OSN can be traced in another OSN.
- At present we are able to predict gender by using the features of the name of the user.
- By the end of this semester project, we expect to have predicted these gender, occupation, and interests with higher accuracy. Besides these attributes, we also expect to predict location of a user based on maximum frequency of their tweets, age based on their interests, etc.

References

[1] C. Matthews, “1 billion people used Facebook on Monday.” Article (CrossRef Link).

[2] R. C. He, “Introducing new Like and Share buttons,” Article (CrossRef Link).

[3] A. Mitchell, J. Kiley, J. Gottfried and E. Guskin, The Role of News on Facebook. Article (CrossRef Link).

[4] T. Risen, Happy Birthday: Facebook Celebrates Its 10th Birthday, Article (CrossRef Link).

[5] F. Lam, K. T. Chen and L. J. Chen, “Involuntary Information Leakage in Social Network Services,” Third International Workshop on Security (IWSEC), Nov. 2008. Article (CrossRef Link).

[6] A. N. Joinson and C. B. Paine, “Self-disclosure,” privacy and the Internet, In the Oxford Handbook of Internet Psychology, pp. 237-252, 2007. Article (CrossRef Link).

[7] K. R. Goldner, “Self Disclosure on Social Networking Websites and Relationship Quality in Late Adolescence,” ETD Collection for Pace University, Jan. 2008. Article (CrossRef Link).

[8] E. McCollister, T. Grance and K. A. Scarfone, “Guide to Protecting the Confidentiality of Personally Identifiable Information (PII),” In: NIST SP - 800-122. pp 58, Apr. 2010. Article (CrossRef Link).

[9] B. Krishnomurthy and C. E. Wills, “On the Leakage of Personally Identifiable Information via Online Social Networks” in Proc. of the 2nd ACM Workshop on Online Social Networks (WOSN), Aug. 2009. Article (CrossRef Link).

[10] *Po-Ching Lin and Pei-Ying Lin*, Unintentional and Involuntary Personal Information Leakage on Facebook from User Interactions