DOI: http://dx.doi.org/10.26483/ijarcs.v8i8.4842

ISSN No. 0976-5697

Volume 8, No. 8, September-October 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A PRIORITY BASED TEXT ENCODING METHOD FOR USER PRIVACY PRESERVATION INSOCIAL STREAMS

M.VijayaMaheswari
PhD Research Scholar,Department of Computer Science
Karpagam Academy of Higher Education
Karpagam University
Coimbatore, Tamilnadu,India

Dr.S.ManjuPriya
Associate Professor, Department of Computer Science
Karpagam Academy of Higher Education
Karpagam University
Coimbatore, Tamilnadu,India

Abstract: Event detection and personal information protection were the emerging important research topics in social media analysis. Privacy preservation is more important to keep the user's personal details to be secured in social media. Social Networks will generate large number of text records in different formats which contains both personnel and other textual information.Privacy preservation in data mining deals with protecting the privacy of individual data or sensitive knowledge of the users of the social networks without sacrificing the utility of the data.Priority based text encoding technique will improves the security of personal information of users. It helps to improve the privacy over clients or reviewer's personal information and make social communication activity secured.

Keywords: social network services, privacy preserving, priority based text encoding, database, anonymization.

1. INTRODUCTION

Event detection and personal information protection were the emerging important research topics in social media analysis[17].
Event detection and privacy preserving are the most important research topics in social media analysis. Social Networks provides large number of text records in various formats which contains both personnel information and reviewers comments[13].The users have started to communicate and share their reviews through social sites which attracted the marketers to extend their business through the online social sites[3].
Data mining for preserving the privacy of the reviewers deals with protecting the privacy of individual data or sensitive knowledge of the reviewers who are all using social networks, without sacrificing the utility of the data[20].
Social Networks will generate large number of text records in different formats which contains both personnel information and other textual information[8].
Privacy preserving in social networks is very much needed to improve the privacy over client's or reviewer's personal information [16]. It also helps to make social communication activity secured and protected [11].

2. LITERATURE SURVEY

- Jiaming Xu et.al., proposed a system for Self-Taught convolutional neural networks for short text clustering. Flexible self taught convolutional neural network framework for short text clustering was proposed in this work.It learns the non biased deep text representation in an unsupervised manner. Determining the window size in convolutional neural network was very difficult task [9].
- Prashant Jawade et.al., proposed a system for confidential database through privacy preserving. He proposed a system for updating the confidential database with preserving the privacy of it. To solve this problem two methods were proposed. They are suppression and generalization based k-anonymous and confidential database. Beside the paper, it is dealing with the case of malicious parties by the introduction of non-colluding third party[7].
- Wen Hua et.al., proposed a method to Understand Short Texts by Harvesting and Analysing Semantic Knowledge. Prototype System Exploiting semantic knowledge provided by the a well known knowledgebase was constructed in this work. Semantic knowledge based harvesting was done in this work which increases the difficulty of prediction rate [2].
- Jemal Abawajy et.al., presented an in-depth survey of the state-of-the-art privacy preserving techniques for social network data publishing, metrics for quantifying the anonymity level provided and information loss as well as challenges and new research directions. The survey helps the readers to understand the threats, various privacy preserving mechanisms and their vulnerabilities to privacy breach attacks in social network data publishing as well as to observe common themes and future directions [1].
- Cedric De Boom Steven Van Canneyt et.al., proposed a model for the Representation learning for very short texts using weighted word embedding aggregation. Due to low dimensional representations, sufficient information was not obtained to improve accuracy of classifier. Semantic word embeddings and frequency information based framework was constructed for arriving at low dimensional representations for short text designed to capture semantic similarity[5].
- Kaziwasif Ahmed et.al., proposed a noble approach of group recommendation preserving the identity of user

from the unauthorized attacker based on the concept of k-anonymity. They tackled the novel and important problem of preserving privacy in personalized community recommendation framework that provides users with the community recommendations while keeping the users preferences hidden from other unauthorized entities based on k-anonymity. This research will help the users to be a member of a community of their own interest without privacy issues [15].

- Zheng Yu et.al., proposed a method for Understanding Short Texts through Semantic Enrichment and Hashing. Encodes the meaning of text into a binary codes.Encode was created by deep neural network which takes a very long time and the network can provide wrong answer under some circumstances[10].
- Dongsheng Li et.al., proposed YANA (short for "You Are Not Alone") a user group based privacy preserving recommender system for users in online social communities. In this system, users are organised into groups with diverse interests and interact with the recommender server via interest – specific pseudo users, so that the individual users personal interest information remains hidden from the server. A suit of secure multiparty computation protocols and recommendation strategies are proposed to protect user privacy from the group members in the recommendation process [19].

3. EXISTING SYSTEM

In Existing system a Semantic word embeddings and frequency information based framework was constructed for arriving at low dimensional representations for short text designed to capture semantic similarity.Due to low dimensional representations, sufficient information was not obtained to improve accuracy of classifier.

A Short Texts through Semantic Enrichment and Hashing technique was used. It encodes the meaning of text into a binary codes. Encode was created by deep neural network which takes a very long time and the network can provide wrong answer under some circumstances .

A Prototype System Exploiting semantic knowledge was provided by the well known knowledgebase for privacy preservation. Semantic knowledge based harvesting increases the difficulty of prediction rate.

A Flexible self taught convolutional neural network framework for short text clustering was used with the non biased deep text representation in an unsupervised manner. Determining the window size in convolutional neural network was very difficult task.

A semi trusted proxy model is employed for data storage activity and it is not fully trusted on securing the sensitive data.In this modelthe encryption, decryption and key generation process will takemore computation time[4].

4. PROPOSED WORK

User's personal details and other data security is provided based on the proposed Priority based Text Encoding algorithm. In this PTE algorithm, the user's personal details are categorized into Email Id , Password , Phone Number and other data including name,address,location etc are anonymized based on the different method of anonymization techniques.

This proposed PTE work create effective Priority based text encoding technique that will improve the security of users / reviewers or personals whose comments are detected.

It helps to improve the privacy over clients or reviewer's personal information and make social communication activity secured [18].

STEPS

- Initially the user /reviewer personal details will be collected from social media along with their comments[25].
- The user reviews/comments are the input, which are the user's feedbacks about the products that are taken from social sites[14]. Input data are divided into user's personal data and user's review data [12].
- The user profile information may include, user name, age, mail id, password, contact number, living residence address, and other location details[21].All these personal information are very sensitive information and seems to be preserved to avoid security issue[22].
- The data are categorized intoEmail Id, Password, Phone Number and other data including name,address,location etc[24].These data are anonymised individuallythrough the proposed Priority Based Text Encoding (PTE) Algorithm.

5. PROPOSED ALGORITHM

Priority Based Text Encoding (PTE)

For user privacy preserving, the details of the user are hidden by applying **Priority Based Text Encoding (PTE)** method. This method will anonymize the personal details of users. It will hide the sensitive information of users in social review analysis.

In this PTE method, the user profile information may include, user name, age, mail id, password, contact number, living residence address, and other location details. All these personal information are very sensitive information and seems to be preserved to avoid security issue.

The data are categorized into four types, Email id, Password, Phone number and other words[23].

Input: Reviewer's Personal Information

Output: Converted into Anonymized of reviewer's personal information

Step-1: Select the key \\ Key→1 or 2 or 3 or 4…
Step-2: temp=0
Step-3: if Sentence (i) ==@,
Step-4: temp=temp+1; end
For Email-id:
Step-5: if temp==1
Step-6: then, Sentence='$\$_1, \$_2, \dots, \$_{key}$'
Step-7: end
For Password:
Step-8: get the length(Password)
Step-9: for i=1 to length(Password)
Step-10: Password(i)= ∗; end for
For Phone Number:
Step-11: for i=1: length(P) \\ P→ Phone Number
Step-12: if P(i) is not equal to space;
Step-13:$P_N = P(i)$;

end if \\ P_N→New Phone Number
Step-14: $A_P = {}^{Temp_2}/_{Temp_1}$
end for
\\ A_P→Anonymize Phone number
Other Words:
Step-15: for i=1 to length(words)
Step-16: W_{ASCII}→ASCII for of words
Step-17: W_A=convert ASCII to word;
\\ W_A→Anoymized Words
end for

6. EXPERIMENTAL ANALYSIS

- The Experimental analysis is carried out on Matlab2009 in a Windows 7 Platform. It is the user reviews or comments about the Amazon kindle product.
- The Priority Based Text Encoding method is used for anonymization. It improves the privacy of the User's personal details.

7. RESULTS AND DISCUSSIONS

The user personal details collected from the social media are specified in the fig 1. The personal detail contains of user / reviewer name, age, company name, location, address, phone number and Email Id.

	Name	Age	Company Name	Loca
1	Graiden	43	Lorem Eu PC	Clarksvi
2	Octavius	63	Aliquet Institute	Whitby
3	Maia	27	Lorem Vitae Odi...	Seborga
4	Simon	25	Dolor Corp.	ChaÃ±a
5	Nayda	36	Ridiculus Incorp...	Thalass
6	Anthony	94	Tellus Phasellus ...	Bareilly
7	Fiona	79	Nunc Ullamcorp...	Fortune
8	Byron	24	Sed Inc.	Guildfor
9	Brooke	62	Neque Vitae LLC	Gavorra
10	Caleb	64	Sed Dictum Proi...	Panjim
11	Brynn	70	Consectetuer A...	Coleville
12	Caesar	78	Ut Ipsum Ac As...	Etobicol

Fig.1: User Personal Details

The reviewer comments are represented in the fig 2. It is the user or reviewer comments about the kindle product [6].

	List of User Reviews
1	Drug Runners and a U.S. Senator have something t...
2	Heres a single, to add, to Kindle. Just read this 19th...
3	If you tire of Non-Fiction.. Check out http://www.a...
4	"Ghost of Round Island" is supposedly nonfiction.
5	Why is Barnes and Nobles version of the Kindle so ...
6	@Maria: Do you mean the Nook? Be careful, books ...
7	kindle is awesome! mines great
8	I love mine!!!!
9	Meh. I think Singles are a bad idea. Big name author...
10	My daugjhter loves hers!
11	I am not sure if i just got my update but now i dont h...

Fig.2: User Reviews / Comments

The user details and comments about the particular product are given as input to processing phase for anonymization is specified in the fig 3 and fig 4.

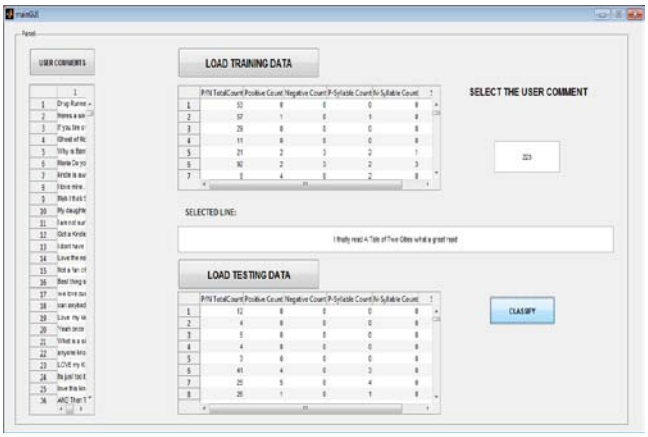

Fig.3: User comments in processing phase

User Details

User Details

	Name	Age	Company Name	Location	Region	Phone No	Mail Id	Password
1	Graiden	43	Lorem Eu PC	Clarksville	CursÃ§ao	01 06 12 13 87	mi@augueutl...	XAL29UEC2JG
2	Octavius	63	Aliquet Institute	Whitby	United States	07 89 03 01 37	turpis.Aliqua...	EBE10JOM9LP
3	Maia	27	Lorem Vitae Odi...	Seborga	Iraq	09 17 43 44 83	Nullam.ut@u...	MMJ18IFT5XR
4	Simon	25	Dolor Corp.	ChaÃ±ara	France	09 38 08 49 55	Nam.ac@dia...	WCV64JEL2...
5	Nayda	36	Ridiculus Incorp...	Thalassery	Ã...land Isla...	07 10 04 47 82	nec.quam@ti...	QBW50PFG6...
6	Anthony	94	Tellus Phasellus...	Bareilly	Anguilla	05 55 27 05 00	vehicula.Pell...	KWR04CJU3...
7	Fiona	79	Nunc Ullamcorp...	Fortune	Iraq	05 09 26 31 31	nunc.In.at@...	LMB50EYH9...
8	Byron	24	Sed Inc.	Guildford	Spain	06 94 23 64 59	et.commodo....	DDH54XNZ6...
9	Brooke	62	Neque Vitae LLC	Gavorrano	Andorra	04 50 11 93 74	eu.nulla@in.ca	LAU85JNO5UF

Fig.4: User personal data without anonymization.

After the preservation of user details , the user personal details are anonymized and it is represented in the fig 5.

Anonimized Details

	Name	Age	Company Name	Location	Region	Phone No	Mail Id	Password
1	Bw\n_j	8	Gtmjh%@zU>	>q\wfxqngq`	>zmf¾~\t	0.0400	$$$$$	**********
2	Jhofqnpx	12	<qdvpjc%Dsnyd...	Rmdyj~	Psdy`iXofojn	0.0400	$$$$$	**********
3	Hfdf	5	Gtmjh%Qnof`%...	Njjtm\	Dw!v	0.0400	$$$$$	**********
4	Nnhti	50	?tgtm%>tmu)	>mE~fmfg	Aw!s^j	0.0400	$$$$$	**********
5	Ifh	7	Mn_n^zgzn%Ds...	Omlq\xnjm~	¾VygfiNnq!s_x	0.0400	$$$$$	**********
6	<somjst	18	OjgqpxUcfnjgqp...	=fmjdqg~	<sbzdqgf	0.0400	$$$$$	**********
7	Anjs\	15	!zihZgq\r*tmu`w...	Atmyps`	Dw!v	0.0400	$$$$$	**********
8	=~mti	4	Nj_%Ds^3	Bzdq_kjw_	Nu!ni	0.0400	$$$$$	**********
9	=wjtfj	12	!Jz`%Qnof`%GQ>	Bfqtmw\sj	<s_tmw\	0.0400	$$$$$	**********

Fig.5: User details after anonymization

The reviewer personal details and comments are collected from the social networks. The collected information contains both numeric and character data. Both the data are anonymized and de-anonymized individually by using Priority Based Text Encoding Algorithm. User sensitive information such as name, age, phone number, email id are hidden by the PTE algorithm. For admins of the particular siteor for other users only the hidden information will be displayed. The reviewer comments will be displayed along with the user hidden sensitive information. Priority Based Text Encoding(PTE)algorithm in privacy preserving hides all the personal and sensitive information of the user. It helps to make the social communication secured.

8. CONCLUSION& FUTURE ENHANCEMENT

The privacy preservation for the user personal details has been implemented with the novel Priority Based Text Encoding model. This modelis implemented for securing the user personal details and sensitive information. The information seems to be secured and preserved to avoid the issue of security of users. By applying this Priority Based Text Encoding (PTE) method, it will anonymized and hide the personal details of users in social networks. And also this PTE model can be sorted through based on the location of users. It will help the users to purchase the products based on location. Further this model can be implemented to extract the online reviews from the other social sites(twitter and etc.) and tested for accuracy. A recommender system can be modelled based on the classification of the reviews and can be implemented in product sale site like flip kart and etc.

REFERENCES

[1] Jemal Abawajy, Mohd Izuan Hafez Ninggal and Tutut Herawan, “Privacy Preserving Social Network Data Publication”, IEEE TRANSACTIONS 08 March 2016.

[2] Hua, Wen, et al. "Understand Short Texts by Harvesting and Analyzing Semantic Knowledge." IEEE transactions on Knowledge and data Engineering 29.3 (2017): 499-512.

[3] Shirakawa, Masumi, et al. "Wikipedia-based semantic similarity measurements for noisy short texts using extended naive bayes." IEEE Transactions on Emerging Topics in Computing 3.2 (2015): 205-219

[4] Xu, Bei, and Hai Zhuge. "An angle-based interest model for text recommendation." Future Generation Computer Systems 64 (2016): 211-226.

[5] De Boom, Cedric, et al. "Representation learning for very short texts using weighted word embedding aggregation." Pattern Recognition Letters 80 (2016): 150-156.

[6] VijayaMaheswari M , Dr.S.ManjuPriya “A Textual Pattern Encoding Method for Privacy Preservation in Social Streams” in Journal of Advanced Research in Dynamical & Control Systems (JARDCS), Issue 5, July 2017,Pages 129 – 134.

[7] Prashant Jawade,Poonam Joshi, “Securing Anonymous and Confidential Database through Privacy Preserving Updates”, International Journal of Applied Information Systems (IJAIS) 2016.

[8] Monika soni , Vishal Srivastva , “Privacy Preserving Data Mining: Comparion of Three Groups and Four Groups Randomized Response Techniques”, International Journal on Recent and Innovation Trends in Computing and Communication(IJRITCC)2013.

[9] Xu, Jiaming, et al. "Self-Taught convolutional neural networks for short text clustering." Neural Networks 88 (2017): 22-31.

[10] Yu, Zheng, et al. "Understanding short texts through semantic enrichment and hashing." IEEE Transactions on Knowledge and Data Engineering 28.2 (2016): 566-579.

[11] Smita R Kapoor, Vismay Jain, R.C.Jain ,” A Privacy Preserving Repository For Data Integration Across Data Sharing Services”, International Journal of Engineering Research & Technology (IJERT) 2013.

[12] M.Vijaya Maheswari, Dr. T.Christopher ,”A Comparative Study on Various Approaches for event Detection in Social Streams”,International Journal of Engineering Research & Technology (IJERT) 2015.

[13] M.Vijaya Maheswari, Dr. T.Christopher ,”A Review on Cluster Based Approach in Data Mining”,International Journal of Engineering Research & Technology (IJERT) 2015.

[14] K.Vidhya.R.Ramapraba,B.Suganya , “ Privacy Preserving for sharing Social Networks Data “ ,International Journal of Advanced Research in Computer and Communication Engineering 2014.

[15] Kaziwasif Ahmed , Israt Jahan Mouri, Rianon Zaman , Nilufa Yeasmin, “A Privacy Preserving Personalized Group Recommendation Framework”, IEEE 6th International conference on Advanced Computing 2016.

[16] Jieming Zhu,Pinjia He, Zibin Zheng,Michael R. Lyu ,”A Privacy-Preserving QoS Prediction Framework for Web Service Recommendation”, IEEE International conference on web services (ICWS) 2015.

[17] Linke Guo,Chi Zhang,Yuguang Fang, “A Trust-Based Privacy-Preserving Friend Recommendation Scheme for Online Social Networks”, IEEE Transactions on Dependable and Secure Computing 2014.

[18] Nikolaos Polatidis, Christos K.Georgiadis, Elias Pimenidis, Haris Mouratidis,” Privacy-preserving collaborative recommendations based on random perturbations”, Elsevier Expert Systems with Applications 2017.

[19] Dong Sheng Li, Qin Lv,LiShang, Ning Gu, “Efficient privacy-preserving content recommendation for online social communities”, Elsevier Neurocomputing 2017.

[20] Tingting Feng, Yuchun Guo, Yishuai Chen,” Can user privacy and recommendation performance be preserved simultaneously?”,Elsevier Computer Communications 2015.

[21] Gautham Pallapa, Sajal K.Das, Mario Di Francesco, Tuomas Aura, “Adaptive and context-aware privacy preservation exploiting user interactions in smart environments”, Elsevier Pervasive and Mobile Computing 2014.

[22] Baozhen Lee, Weiguo Fan, Anna C.Squicciarini, Shilun Ge, Yun Huang,” The relativity of privacy preservation based on social tagging”, Elsevier Information Sciences 2014.

[23] Alper Bilge,Ihsan Gunes,Huseyin Polat,”Robustness analysis of privacy-preserving model-based recommendation schemes”, Elsevier Expert systems with Applications 2014.

[24] Lifang Zhang, Zheng Yan, Raimo Kantola,” Privacy-preserving trust management for unwanted traffic control”, Elsevier Future Generation Computer Systems 2016.

[25] Gary Blosser, Justin Zhan,”Privacy preserving Collaborative Social Network”, IEEE International Conference on Information Security and Assurance 2008.