ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

Generating String Recommendation Efficiently and Privately

Neelima Ramnath Satpute[1] and Prof. Hyder Ali Hingoliwala[2]

1. ME Scholar, Department of Computer Engineering, JSPM'S JSCOE, Pune University, Maharashtra, India.

2. Professor, Department of Computer Engineering, JSPM'S JSCOE, Pune University, Maharashtra, India.

Manuscript Info

Manuscript History:

Received: 14 April 2015
Final Accepted: 12 May 2015
Published Online: June 2015

Key words:

Homomorphism encryption, privacy, steaming, stop word, recommender systems, secure multiparty computation.

Corresponding Author

Neelima Ramnath Satpute

Abstract

Recommender systems have turned into a critical tool for personalization of online services. Producing proposals in online administrations relies on upon privacy-sensitive data gathered from the clients. Conventional information assurance mechanisms concentrate on access control and secure transmission, which give security just against malicious third parties, yet not the service provider. This makes a genuine security hazard for the clients. In this paper, we expect to ensure the private information against the service provider while saving the usefulness of the framework. We propose encoding private information and preparing them under encryption to create suggestions. By presenting a semi-trusted third party and utilizing data packing, we develop an exceedingly proficient framework that does not require the dynamic interest of the client. We additionally exhibit a comparison protocol, which is the first to the best of our insight that analyzes different qualities that are packed in one encryption. Directed trials demonstrate that this work opens a way to generate private recommendations in a privacy-preserving manner. The existing system work on only the integer recommendation but in our propose work we implement on phrase and string recommendation by applying steaming and stop word removal.

INTRODUCTION

A large number of individuals are utilizing online administrations for different every day exercises, a considerable lot of which oblige imparting individual data to the administration supplier. Consider the accompanying online administrations:

Social Networks:

People use social networks to contact other people, and make and offer content that incorporates individual data, pictures, and features. The administration suppliers have admittance to the substance gave by their clients and have the privilege to process gathered information and appropriate them to third parties. An extremely regular administration gave in social networks is to produce suggestions for discovering new companions, gatherings, and events using collaborative filtering techniques. The information needed for the collaborative filtering algorithm is gathered from different assets including clients' profiles and behaviors.

Online Shopping:

Online shopping services improve the probability of a buy by giving customized recommendations to their clients. To discover administrations and items suitable to a specific client, the administration supplier procedures gathered client information like client inclination and click logs.

IP-TV:

A set-top box with high storage capacity and transforming power takes its place almost in every family unit. The administration supplier's utilization brilliant applications screen individuals' activities to get (statistical) data on people's viewing habits their preferences and abhorrence. In light of the data gathered from the clients, the administration supplier prescribes customized computerized substance like TV programs, movies, and items that a specific client may discover fascinating.

In the greater part of the above services, recommender frameworks in view of collaborative filtering techniques that gather and methodology individual client information constitute a fundamental piece of the administration. On one hand, people advantage from online services. Then again, direct access to private information by the administration supplier has potential privacy risks for the clients since the information can be prepared for different purposes, exchanged to outsiders without client assent, or even stolen. Recent studies demonstrate that the protection contemplations in online services appear to be a standout amongst the most essential variables that undermine the sound development of e-business. Thusly, it is essential to secure the security of the clients of online administrations for the profit of both people and business.

Recommender Systems

The objective of a recommender framework is to deliver convincing recommendations to a social event of clients for things or items that may captivate them. Recommendations for books on Amazon, or films on Netflix, are veritable outlines of the operation of industry quality recommender frameworks. The setup of such recommendation engines depends on upon the space and the particular traits of the information available. For example, film watchers on Netflix chronically give appraisals on a size of 1 to 5. Such an information source records the way of associations amidst clients and things.

Also, the framework may have induction to customer particular and thing particular profile attributes, for instance, demographics and item portrayals, separately. Recommender frameworks differ in the way they inspect these information sources to make ideas of preferring amidst clients and things, which can be used to perceive by and large matched sets. Collaborative Filtering systems research legitimate associations alone, while Content-based Filtering frameworks are in light of profile properties; and crossover methodologies try to combine both of these diagrams. Recommendation engines are the product that prescribes what we should watch or read or listen to next. They help us deal with the immense numerous choices the Web offers.

Security

Recommender systems require two types of trust from their users. First, since the recommender must receive substantial information about the users in order to understand them well enough to make effective recommendations, they must trust that the system will protect their information appropriately. Second, automated recommender systems are often fairly opaque to their users. Although the algorithms used are easy to understand in principle, a user is usually not presented with sufficient information to know exactly how or why an item is being recommended to her. Thus, in order for a recommendation to be accepted, the user must trust that the recommendations are accurate.

In this paper we discuss about the related work, the proposed method where we design our solution, mathematical model, algorithms, and system architecture, the Implementation details and proved the simulation result with graph and at last conclusion in given.

Material and Methods

related work

In the last few decades, necessitate for providing the privacy to the online services, mainly those using the mutual filtering techniques. Number of researchers work for providing security to the online research system, in this section we discussed the work done by the researchers.

In [1], R. Agrawal and R. Srikant tackle some questions like the key task in the data mining is the expansion of models regarding aggregated data, is it easy for developing precise model without access to specific information in individual data records? Author considered the real case of building a decision tree classifier from the training data in which the values of individual records have been troubled. They proposed architecture for reconstruction procedure for accurately estimate original data values. By using these reconstructed distributions anyone can be able for building the classifier whose accuracy is comparable to the accuracy of classifier built with the original data.

In [2], Y. Lindell and B. Pinkas introduced the idea of privacy preserving in data mining. In this model two parties' outstanding secret databases which wish to run a data mining protocol on the base of their databases, without revealing any unnecessary information. They present a solution which is significantly more competent than the generic solution. In this method each party perform by itself a computation of the same order as compute the ID3 algorithm for its own database.

In [3], S. Zhang, J. Ford, and F. Makedon represent propose two information recreation strategies that get unique private data from disguised information in existing perturbation collective filtering schemes. One strategy is based on k-means clustering and other uses single value decomposition (SVD). Author has conducted hypothetical and experimental analysis examination on the difference between unique information and reproduced information. Their tests demonstrate that both strategies can infer an impressive measure of unique data. This study serves to focus an exact exchange
between proposal precision and client protection in irritation plans.

In [4], J. F. Canny propose a substitute scheme in which clients control all their log control. They illustrate a protocol whereby a group of clients can process public "aggregate" of their information which does not uncover individual clients' information. The aggregate permits customized recommendation to be figured by individuals from the community, or by outsiders. The numerical algorithm is fast, powerful and accurate. This technique reduces the community filtering task to an iterative computation of the aggregate requiring only expansion of vectors of client information. At that point we utilize homomorphic encryption to permit sums of encrypted vectors to be processed and decrypted without uncovering individual information. They give verification scheme.

for all parties in the calculation. This framework can be executed with untrusted servers or with additionally infrastructures as a fully peer to peer system.

In [5], J. F. Canny represent another technique for collaborative filtering which secures the protection of individual information. The technique is based on the probabilistic element investigation model. Privacy protection is given by a distributed protocol which is described somewhere else, however plot in this method. The factor analysis approach handles missing information without requiring default values for them.

In [6], Z. Erkin, M. Beye, T. Veugen, and R. L. Lagendijk proposed the method to encrypt the security sensitive information and produce recommendations by handling them under encryption. With this approach, the

administration supplier adapts no data on any client's inclination or the proposals made. The proposed strategy is based on homomorphic encryption scheme and secures multiparty computation (MPC) procedures. The overhead of working in the encrypted domain is minimized by packing data as shown in the complexity analysis.

In [7], Z. Erkin, M. Beye, T. Veugen and, R. L. Lagendijk proposed technical systems to secure the security of individual in a recommender framework. The proposition is established on homomorphic encryption, which is utilized to obscure the private rating data of the clients from the service provider. While the client's security is regarded by the service provider, by producing recommendations using encrypted client evaluations, the service supplier's commercially important item are ensured against inquisitive elements, in turn. This method investigates straightforward and effective cryptographic strategies to create private recommendation using a server-customer model, which neither relies on trusted party, nor requires interaction with per user.

In [8], Z. Erkin, M. Beye1, T. Veugen and R. L. Lagendijk propose technological components to secure the protection of individual in a recommender framework. This method is established on homomorphic encryption, which is utilized to obscure the private rating data of the clients from the service provider. While the client's protection is respected by the service provider, by creating recommendations using encrypted customer rating, the service supplier's commercially valuable thing similarities are protected against curious entities, in turn. This proposition investigates basic furthermore productive cryptographic systems to create private proposals using a server-customer model, which not one or the other depends on (trusted) outsiders, nor obliges connection with associate clients.

Proposed SYSTEM

In the proposed System we are using the sentences or words or phrases for generating recommendation. First we are stemming the sentence or phrase after we are applying stop word removal algorithm so that the coming result after stopword removal is compares with vector of synonyms to get the actual recommendation. We are proposed the system which is generating recommendation on text rating.

The Proposed System generates the Recommendation for the Product based on review in the String form. Here we are using the three agents like User, Privacy Service Provider and Service Provider For the security Purpose we are generation the recommendation on encrypted data. For The Security we are using homomorphic Encryption algorithm i.e. paillier algorithm For Encryption and Decryption and PSP is responsible for decryption of generated Recommendation:

System Architecture

Mathematical Model

1. *The Paillier scheme*

$$\varepsilon_{pk}(m,r) = g^m \cdot r^n \bmod n^2$$

Where,
$n \rightarrow$ product of two large prime numbers *p, q, g* generates a subgroup of order *n* and *r* is a random number in Z_n^*
$(n, g) \rightarrow$ Public Key
$(p,q) \rightarrow$ Private Key
$Z_n \rightarrow$ Message space
$Z_{n^2}^* \rightarrow$ Cipher text space

The homomorphic property of the Paillier cryptosystem can be easily verified as shown below:

$$\varepsilon_{pk}(m_1,r_1) \times \varepsilon_{pk}(m_2,r_2) = \varepsilon_{pk}(m_1 + m_2, r_1 \cdot r_2)$$

$= g^{m_1} \cdot r_1^n \times g^{m_2} \cdot r_2^n \bmod n^2$
$= g^{m_1+m_2} \cdot (r_1 \cdot r_2)^2 \bmod n^2$

2. Collaborative Filtering

For two users, *A* and *B*, with preference vectors $V_A = (v_{(A,0)}, \ldots, v_{(A,M-1)})^T$ and, where *M* is the number of items and $v_{i,j}$ is a small, positive integer, the Cosine similarity is given by

$$sim_{(A,B)} = \frac{\sum_{m=0}^{M-1}(v_{(A,m)} \cdot v_{(B,m)})}{\sqrt{\sum_{m=0}^{M-1} v_{(A,m)}^2 \cdot \sum_{m=0}^{M-1} v_{(B,m)}^2}}$$

$$= \sum_{m=0}^{M-1} \frac{v_{(A,m)}}{\sqrt{\sum_{m=0}^{M-1} v_{(A,m)}^2}} \cdot \frac{v_{(B,m)}}{\sqrt{\sum_{m=0}^{M-1} v_{(B,m)}^2}}$$

$$= \sum_{m=0}^{M-1} \tilde{v}_{(A,m)} \cdot \tilde{v}_{(B,m)}$$

$\tilde{v}_{(A,m)}$ and $\tilde{v}_{(B,m)}$ are the normalized vector elements of users *A* and *B*, respectively.

Proposed System Algorithm

- Stemming and Stop word Removal approach:

1: Read a line Entered in the review box
2: Extract words (from this point we called it as token) from the line, clean the token, that is remove punctuation marker attached with token if there is one.
3: Check the dictionary. If a dictionary entry matches with the token, mark token as root word and exit otherwise execute the next step.
4: Look up suffix-list generated manually from the end of the token. If there is a match with the suffix-list Extract and exit.
5: After Removal of stemming the present tokens is matched with stop word list.
6. Remove stop-word found
7. Final remained Tokens.

- *Recommendation Generation algorithm:*

For User:
Step 1: if(user Exist)
Step 2: Select Product From product List
Step 3: Enter Review For the Product
Step 4: Apply StopWord algorithm
Step 5 : Apply Stemming algorithm
Step 6: Store Review In database.
Step 7. Get Public Key From PSP
Step 8: Encryption of all the review in the database by using received Public Key
Step 9: Send Encrypted Review to SP
Step 10: Receive Encrypted Recommendation from Sp
Step 11: Add Random No in the Recommendation
Step 12: Send Result for the Decryption to PSP
Step 13: Receive Final Decrypted Recommendation From PSP
Step14: Remove Random no added in Recommendation
Step 15: Final Generated Recommendation

At Service Provide (SP)

Step 1: Receive Encrypted Review

Step 2: Generate Recommendation for Encrypted review
Step 3: Send Generated Result to User

At Privacy Service Provider(PSP)

Step 1 : Send Public key to User for Encryption
Step 2: Receive Encrypted Recommendation For Decryption
Step 3: Decrypt Recommendation
Step 4: Send Decrypted Recommendation to User

IMPLEMENTATION DETAILS

Hardware Requirement

- **Hard disk : 80 GB**
- **RAM : 512 MB**
- **Processor : Intel Pentium4 or above**

Software Requirements

- Operating System : Independence of Operating System
- Application Libraries: Java and J2EE
- Language : Java
- Front End : Jung tool

Result and Discussion

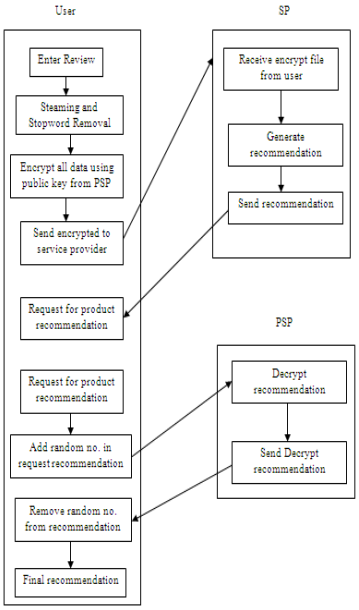

Fig. 1 System Architecture

SIMULATION RESULT

To evaluate situation-specific performance in

Screen Shots

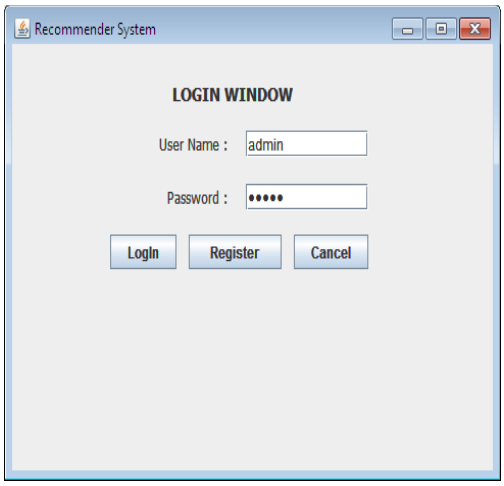

Fig2. Login Form

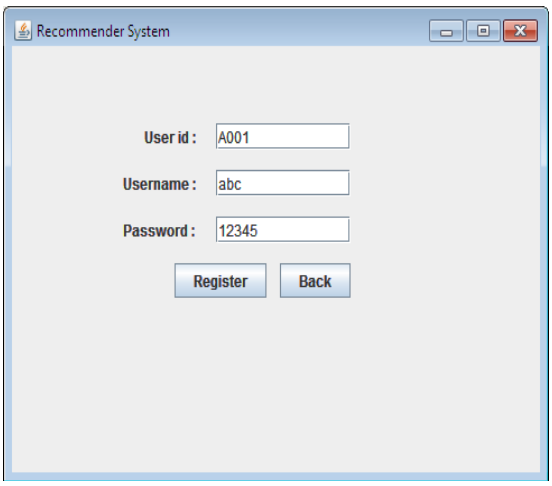

Fig.3 Sign up

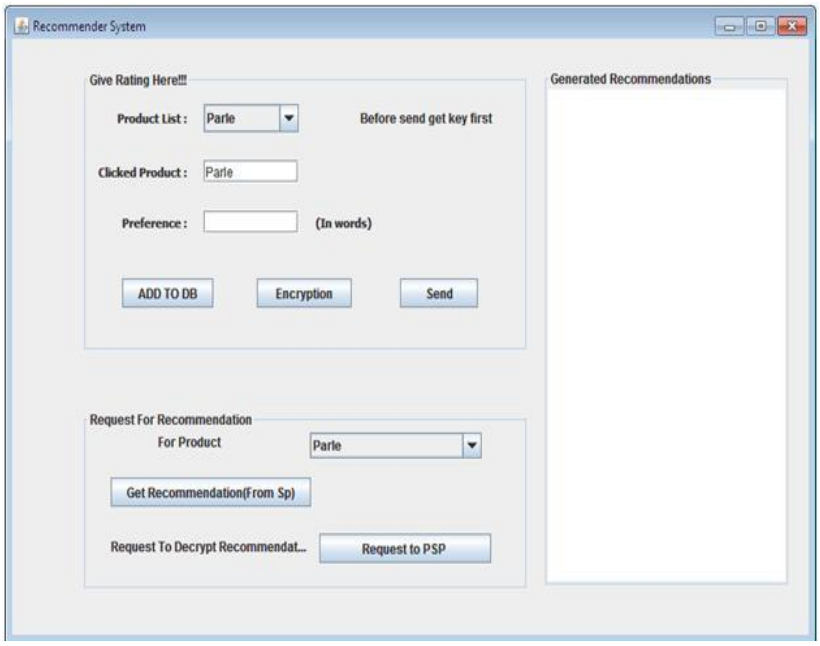

Fig.4 Recommendation System

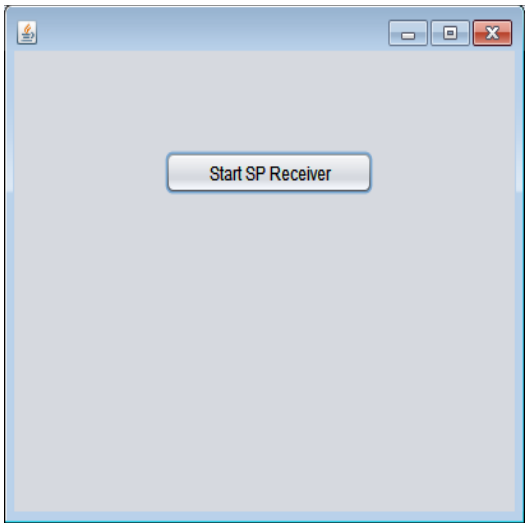

Fig.5 At SP side

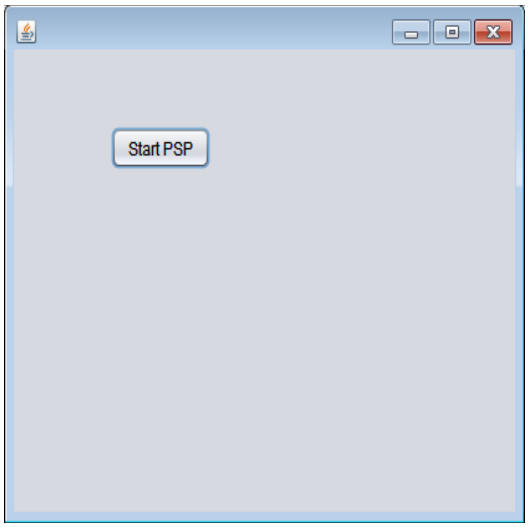

Fig.6 At PSP side

Graphs

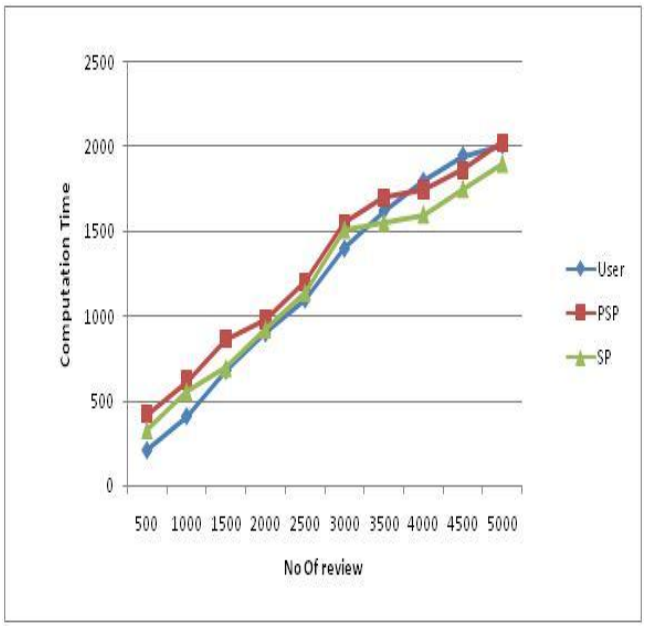

Fig.7 Computational time Vs. No. of Review

This is the Computation Time Graph which Shows the Computation time for three modules we are using for generating recommendation. The Computation time is depending upon the no of review taken for generating the recommendation or the dataset size we used.

Conclusion

With the huge increment in the online administrations gave in last few years; the recommendation systems are expected to be more exact and proficient. The security is the greatest angle for the achievement of any such framework, as the clients don't need their data to be taken care of by the third individual, who may abuse it. The

cryptographic protocol can be utilized for encrypting the client information that encoded information can be transformed by distinctive proposal era system. The suggestion created ought not to endure tradeoff between accuracy and privacy. Generating recommendation using string data reviews will be more users friendly. The recommendation created will likewise as texts and strings which is more helpful for client to understand exact recommendation.

References

1. **R. Agrawal and R. Srikant**, Privacy-preserving data mining, in Proc. SIGMOD Rec., May 2000, vol. 29, pp. 439–450.

2. **Y. Lindell and B. Pinkas**, Privacy preserving data mining, J. Cryptol., pp. 36–54, 2000, Springer-Verlag.

3. **S. Zhang**, J. Ford, and F. Makedon, Deriving private information from randomly perturbed ratings, in Proc. Sixth SIAM Int. Conf. Data Mining, 2006, pp. 59–69.

4. **J. F. Canny**, Collaborative filtering with privacy., in IEEE Symp. Security and Privacy, 2002, pp. 45–57.

5. **J. F. Canny**, Collaborative filtering with privacy via factor analysis, in SIGIR. New York, NY: ACM Press, 2002, pp. 238–245.

6. Z. **Erkin, M. Beye, T. Veugen, andR. L. Lagendijk**, Privacy enhanced recommender system, in Proc. Thirty-First Symp. Information Theory in the Benelux, Rotterdam, 2010, pp. 35–42.

7. **Z. Erkin, M. Beye, T. Veugen, and R. L. Lagendijk**, Efficiently computing private recommendations, in Proc. Int. Conf. Acoustic, Speech