ISSN:0975-9646

SMS Encryption

Rashmi P. Sarode*, Neeraj Manglani

Department of Computer Science

Jagan Nath University, Jaipur, India

***Abstract*— In the application of Computer Science, the performance of classification is generally improved by using supervision in the form of pair wise (must-link and cannot-link) constraints. This paper introduces a rigorous Bayesian framework for semi-supervised clustering which incorporates human supervision in the form of pair wise constraints both in the expectation step and maximization step of the Naive-Bayes algorithm and RSA algorithm. In this framework, a system has been developed for answering all the queries about railways in an elegant way as compared to the existing system which provides information about railways with limited capabilities and is a web based service.**

Keywords— **Classification, RSA algorithm, Naïve-Bayes algorithm.**

I. INTRODUCTION

Information retrieval is the process of collecting information from information resources which are relevant. One can create searches which are based on metadata or on full-text (or other content-based) indexing. This process begins when a user enters queries into the system which are formal statements of information needs, for example search strings in web search engines [1]. In order to answer all the queries about railways, we have developed a system which does not require internet connection. This system will enable any person to enquire about timings, fare, types of train and food availability on the train and it will be controlled by the railway database administrator.

II. LITERATURE REVIEW

A. *NAIVE- BAYES CLASSIFICATION*

This Classification is named after Thomas Bayes (1702-1761), who proposed the Bayes' Theorem [2]. This is a text based clustering. Naive-Bayes classification represents a supervised learning method as well as a statistical method for classification. It calculates explicit probabilities for input query.

This algorithm trains the database such that it will retrieve appropriate answer from database for unknown input which is new to the database. It is a simple classifier with independent and strong assumption. For example, we can consider a fruit to be an apple if it is round, red and 3" in diameter. The Naive-Bayes classifier ponders these features to contribute independently to the probability that this fruit is an apple, irrespective of the presence or absence of the other features [3].

B. *Parameters of NAIVE BAYES*

Suppose the training data contains a continuous attribute x, we first segment the data by the class, and then compute the mean and variance of x in each class. Let μ_c be the mean of the values in x associated with class c and let σ_c^2 be the variance of the values in x associated in class c. Then, the probability of some value given a class, $P(x= v|c)$, can be computed by plugging v into the equation for a Normal distribution parameterized by μ_c and σ_c^2 i.e. [4],

$$P(x = v \mid c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}}$$

C. *RSA ALGORITHM*

One example of asymmetric algorithm is RSA algorithm. It was developed by Ron Rivest, Adi Shamir and Leonard Adelman in 1977 as the first major asymmetric key cryptography algorithm. The name RSA comes from the surnames of these three above research scientists. Ron Rivest was a professor working in Massachusetts Institute of Technology, USA (MIT). He engaged Shamir and Adelman to work on the notion of asymmetric key cryptography [5].

RSA requires keys of at least 1024 bits for security but keys of size 2048 are best suited for security purpose. It is widely used as secure communication channel and for authentication of service provider [6]. RSA is too slow for encrypting large amounts of data and is widely used for key distribution. The approach used here is based on asymmetric algorithm which involves a key pair, a public key and a private key [7].

To communicate securely over any network, one needs to publish the public key. All these public keys are stored in a database which any one can refer to. But the private keys remain with the respective individuals only. It is a very challenging task to create private key from the public key, so RSA is a very prevalent choice in data encryption [8].

RSA also uses the technique of digital signatures, so we know the message is coming from a particular user. The use of digital signature also prevents the message to be altered in the transit. In RSA algorithm we need the keys same as number of participants so this algorithm scales up quite well. There is no problem of key agreement or key exchange here [7].

D. Parameters of RSA

Consider the equation for encryption and decryption as follows:

$CT=PT^E \bmod N$ and $PT = CT^D \bmod N$

Here:

CT and PT are Cipher text and Plain text respectively

E is the encryption key

D is the Decryption key

N is the product of the two chosen prime numbers for factorization [5].

II. EXISTING SYSTEM

The traditional approach to retrieve information of any railway enquiry is restricted with limited capabilities and some information is based on internet.

There are basically two ways to get to know about any railway enquiry:

I. Personal visit to railway station

II. Website of railways.

In the first approach, person will visit the railway station and will gather information of that particular train which he wants to travel. It is a one to one way to retrieve information. But this is manually done process which may contain many flaws such as the ticket counter person will not give exact information which may create many misunderstandings. So this type of approach to retrieve information may lead to disaster.

The second approach is to visit the web site of railways and get information about it. Information retrieval systems based on internet are used to reduce what has been called "information overload". Many websites use Information Retrieval systems to provide access to train timings, fare, type of train and food availability on train. Web search engines are the most visible Information Retrieval Applications. The Architecture of the existing Indian Railway System is shown in Fig 1.

Disadvantages

I. Websites are not updated frequently.

II. More processing time.

III. Need of registration.

IV. Internet connection required.

This is why in our system; we propose to make information retrieval by SMS facility which does not require internet connection. This system will help us to keep track on all the current and upcoming events just by one SMS. Processing time of this system is less than 2 seconds with ease of processing and no need of registration.

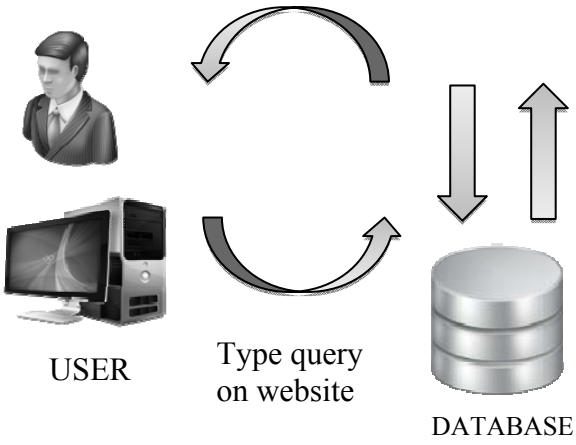


Fig 1.: Architecture of the Existing Railways System

III. DESIGN AND IMPLEMENTATION

A. SYSTEM DESIGN

The client sends a query in SMS to the registered SMS server, the SMS server then communicates with the database thus authenticating the user. The SMS server also categorizes the keywords and analyzes the query. The database generates a response for the SMS server and SMS server sends the response to the client. This architecture can be seen in Fig. 2.1.

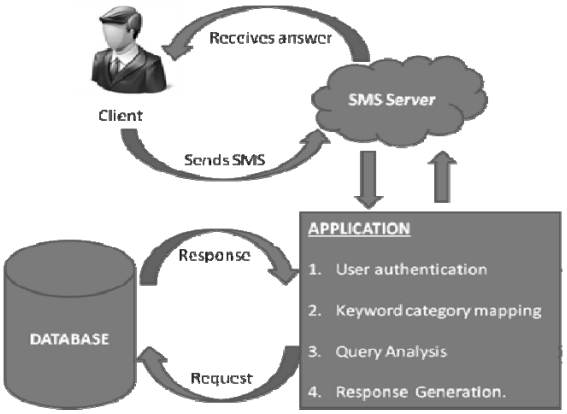


Fig.2.1: ARCHITECTURE of Information Retrieval by SMS

B. Detailed Data Flow Diagram

The query is sent to the application which first groups it into categories by clustering method. Then the application analyzes the SMS query from the clustering and generates a response. A detailed Data Flow Diagram (DFD) of this design is shown in Fig.2.2.

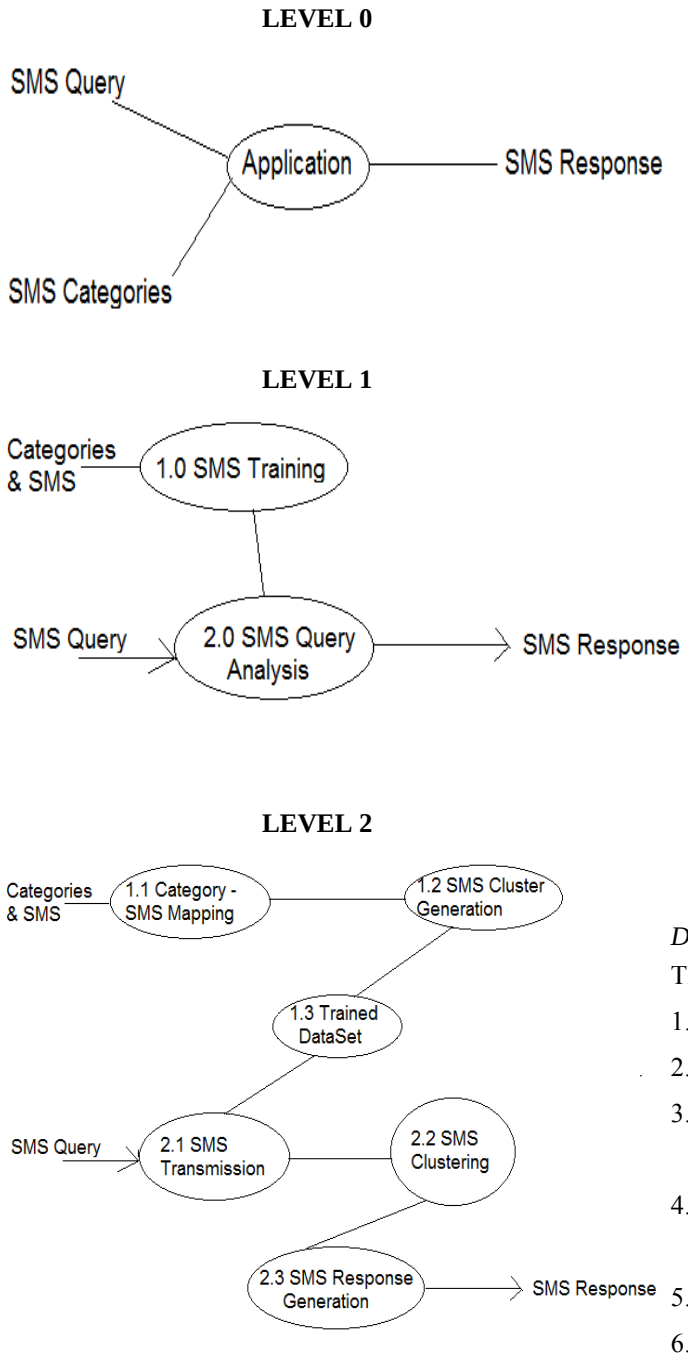


Fig. 2.2: Data Flow Diagram

C. *IMPLEMENTATION*

We introduce a rigorous Bayesian framework for semi-supervised classification which incorporates human supervision in the form of pair wise constraints both in the expectation step and maximization step of the Naive-Bayes Classification algorithm. The encryption and decryption processes of the messages are handled by RSA algorithm. The implementation is illustrated clearly in Figure 3.

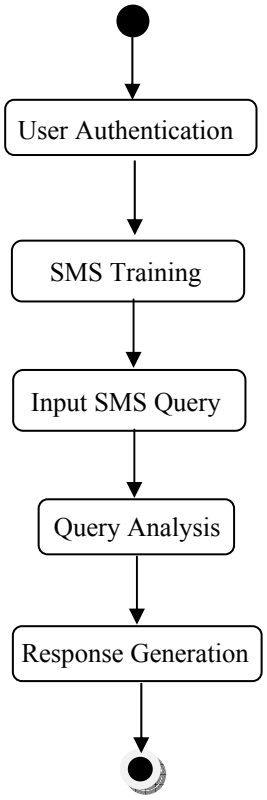


Fig.3: Activity Diagram

D. *METHODS TO BE IMPLEMENTED*

The process requires following steps:

1. Creation of User Interface in JSP using NetBeans.
2. An event handler function is created.
3. All the push buttons with their respective events are created.
4. Training and Testing Databases are created and loaded using MySQL.
5. The query to be classified is selected.
6. Naive-Bayes Classification algorithm is applied and classified.

The following PUSH BUTTONS are used:

- **Sign in:** Retrieve your Username and Password.
- **Register:** Sign up as a new user.
- **Classify:** Click to get Answer of your Query.
- **Home:** Redirect to the home page.
- **Message:** Shows message categories.
- **Logout:** User will log out.

IV. TESTING AND DEPLOYMENT

In Graphic User Interface (GUI) Testing, the GUI is opened without any errors and when buttons are validated on GUI, the buttons are checked one after the other. Then Sign In button allowed registered user to enter into a Home page to type a query and if user is not registered it did not allow that user to go to Home page.

The Home button redirected the user to home page. The Message button worked properly to display the desired categories. The Logout button successfully logged out the user and redirected the user to Login Page.

In Functional Testing, after clicking the register button, the user was successfully registered. After registration, the user typed query into text field which was displayed on home page and the classify button was clicked which returned a new page with correct answer. The Message button displayed the list of available categories. When the classify button was clicked with an already processed query, the same and accurate answer was displayed.

V. RESULTS

The client first registers on the application and then can sign in with his credentials. The login screen is shown in Fig. 4.1. Once the user logs in, he can type any query. The query page is shown in Fig 4.2. Once the user clicks on the classify button, the query is processed and an response is generated through the application which is nearly accurate as shown in Fig 4.3

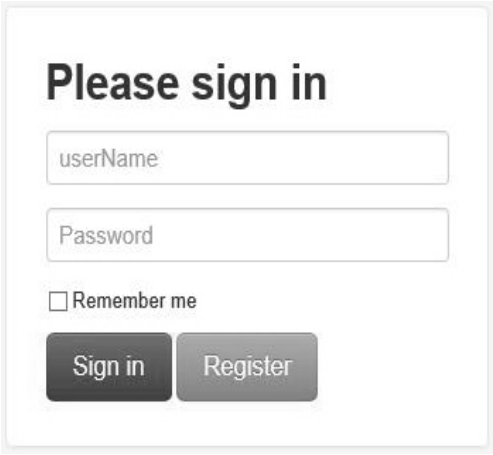


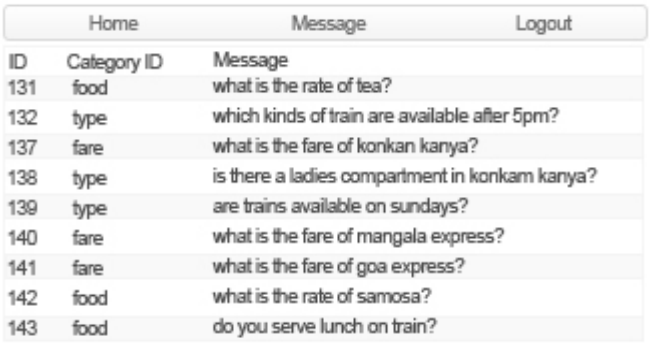
Fig. 4.1 Login Page

Fig. 4.2 Home page with query

Fig. 4.3 Output with an accurate answer

Fig. 4.4 All available categories

We can see all the query categories available if the category button is clicked as shown in Fig 4.4. Then any category can be browsed for the category specific queries as listed in Fig 4.5.



ID	Category ID	Message
131	food	what is the rate of tea?
132	type	which kinds of train are available after 5pm?
137	fare	what is the fare of konkan kanya?
138	type	is there a ladies compartment in konkam kanya?
139	type	are trains available on sundays?
140	fare	what is the fare of mangala express?
141	fare	what is the fare of goa express?
142	food	what is the rate of samosa?
143	food	do you serve lunch on train?

Fig. 4.5 All expected queries

This application can also be tested on any ordinary mobile phone. We just have to send the query to the Registered SMS server. This is shown in Fig 4.6. A few responses received as reply are shown in Fig 4.7.

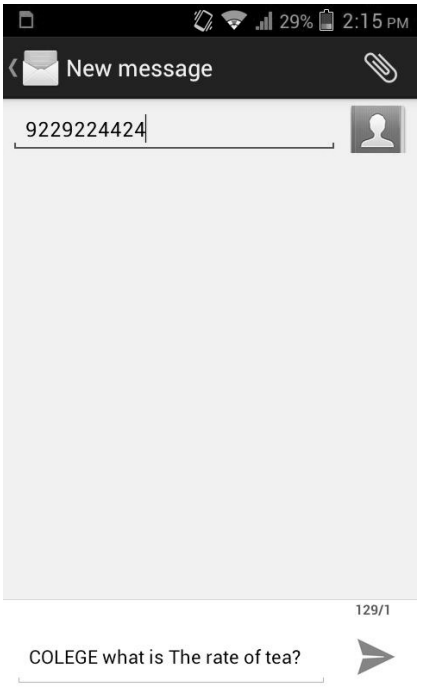


Fig. 4.6 Query shown in mobile

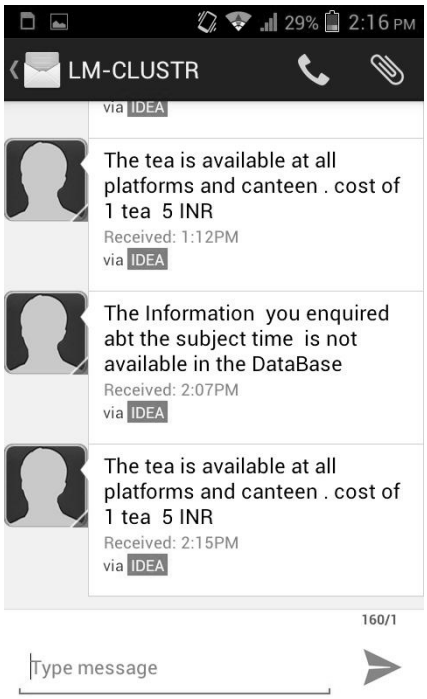


Fig. 4.7 Response by Registered SMS server

VI. CONCLUSION

In this paper, we have been able to develop an interactive system for acquiring information from railways through queries raised by the user on any ordinary mobile phone using simple SMS facility. This system reduces the efforts for gaining the desired information in a shortest time without the use of internet.

VII. FUTURE SCOPE

Once the user's query is classified by Naive Bayes Algorithm and encrypted through RSA Algorithm, it can be further encrypted by Triple DES (Data Encryption Standard). Triple DES is the common name for Triple Data Encryption Algorithm symmetric-key block cipher. It applies the Data Encryption Standard cipher algorithm three times to each data block on the query. Thus, it makes the system much more secure.

Also this concept can be extended to some other level of fields like educational system or hospital system where any student or patient will be able to access any desired and detailed information about the system through a local message. As the scope of the work or the reach of it is large, the idea can be used at any level of hierarchy.

ACKNOWLEDGEMENTS

I would like to thank Dr. Meenu Dave, Dr. Sandeep Poonia and Dr. Renu Bagoria for their constant encouragement and support.

REFERENCES

[1] Information retrieval in Wikipedia, March 27, 2014, from https://en.wikipedia.org/wiki/Information_retrieval
[2] Maharaj E. A. Cluster of Time Series [J], Journal of Classification 2000, 17(2):297-314.
[3] Johnson R.， Wichern D., Applied Multivariate Statistical Analysis (6th Ed) 2007.
[4] Ahmed S., Singha A., Rana M., Mollah N., "Bayesian Approach for Prediction of Interface and Non-interface Residuesfrom Protein Sequence", International Conference on Materials, Electronics & Information Engineering, ICMEIE-2015, 05-06 June, 2015
[5] Kahate, A., "Cryptography and Network Security", 2nd Edn,Tata McGraw-Hill ,2005.
[6] Padmavathi, B., Ranjitha Kumari. S., "A Survey on Performance Analysis of DES, AES and RSA Algorithm along with LSB Substitution Technique", International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064 Volume 2 Issue 4 April 2013 pp 170-174
[7] Ambedkar,B,Bedi, S., "A New Factorization Method to Factorize RSA Public Key Encryption" IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 1, November 2011 pp 242-247
[8] Rangarajan S., Ram N., Krishna V.,"Securing SMS using Cryptography" , International Journal of Computer Science and Information Technologies, Vol. 4 (2) , 2013, 285 – 288