IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-6, June 2018 E-ISSN: 2347-2693

Protecting the Users Information in Personalized Recommendation

P.B. Varpe[1*], M.A. Wakchaure[2]

[1] Computer Department, AVCOE Sangamner, Savitribai Phule Pune University, Pune, India
[2] Computer Department, AVCOE Sangamner, Savitribai Phule Pune University, Pune, India

[*]*Corresponding Author: parivarpe01@gmail.com, Tel.: 7219491110*

Available online at: www.ijcseonline.org

Accepted: 11/Jun/2018, Published: 30/Jun/2018

Abstract— As online purchase has growing nowadays, recommendation becomes important field for today. Due to the regard of privacy, user's unwillingness to expose their private data has become considerable obstacle for the growth of customized recommendation system. So the motive is to safeguard the user's private data. In this work, it is proposed to formulate the dummy preferences set to protect user's sensitive subjects. Firstly, a client based structure for user security assurance is introduced, which does not need any modification to existing algorithms, as well as no trade off to the proposal exactness. Then a privacy protection model formulated by the prime requirements such as similarity in the feature distribution and the degree of exposure is put forth. Feature distribution measures the success of dummy preference profile to envelop actual user profile and the degree of exposure measures the favorable result of dummy preferences to envelop sensitive subject. Finally the implementation algorithm is introduced to meet the actual privacy goal. Proposed system also aims to provide the sentiment analysis of the reviews for the products in order to help the people to identify the good products among the huge number of products available.

Keywords— Personalized Recommendation, Individual Privacy, sensitive subjects, Feature Distribution, Dummy Preferences.

I. INTRODUCTION

With the growing ubiquity of access to online data sources, the recommender frameworks have risen as a capable instrument to lessen data over-burden and give personalized data access for the targeted audience. Recommender frameworks are information filtering frameworks related to different application spaces or sites. They endeavour to fulfil the client's need by giving custom fitted administrations by considering their tastes and attraction. In most cases, these frameworks utilize computational techniques to break down clients past activities and choices. Also, client's related data is utilized for creating the valuable customized suggestion. Recommender frameworks are utilized as a part of different application spaces beginning from social networking sites, e-commerce to online content streaming sites. They are intended to enhance the client experience via consequently separating the broad information about client preferences, practices and giving stuff important to particular clients. Along these lines, recommender frameworks can diminish singular client's intellectual load, and at the same time furnishes them with more significant and important item and administrations. Regardless of the developing fame, these recommender frameworks are not 100% dependable, as the individual data utilized as a part of these frameworks offer ascent to genuine security concerns. Clients whose protection is attacked in any event once are distrustful of utilizing such frameworks in later circumstances.

Recommender frameworks proactively tailor the online items and services as per client's choices and requirements. This procedure of tailoring item and services is known as personalization. Personalization-based framework upgrades the client involvement in numerous ways on the web but also raises the worry for client security. The majority of the recommender frameworks go for giving customized benefit and consequently goes under the personalization-based frameworks classification. For example, MovieLens is a customized recommender framework. This recommender framework proposes film for clients in light of their past observed movies and their feedback. Thus, it is necessary for such framework to think about the choices of its clients before giving the customized recommendation. Amazon.com, a pioneer in the field of web based business, utilizes automated collaborative filtering methods for giving exceptionally customized involvement to clients in view of client's buy history. Client's data as movie rating or buying history prompts better personalization yet additionally contributes in attacking client protection. The protection worries in

collaborative filtering frameworks are high where the framework endeavours to augment the usage of the client's given substance.

Rest of the paper is organized as follows, Section I contains the introduction of proposed framework, Section II contains the related work ,Section III contain the System Architecture and overview, Section IV contain the actual experimental details section V explain Result analysis Section VI concludes research work with future directions

II. RELATED WORK

The Framework in [1] is the system in which they proposed easy but successful privacy preserving structure for QoS-based Web service suggestion. In particular, clients are empowered to obfuscate their private information by data randomization systems before they open the information to a recommender framework. Along these lines, the recommender framework can just gather obfuscate QoS information from clients, and subsequently decrease the hazard to reveal client's privacy. Their privacy-preserving structure is general and can be applied to both the neighborhood-based collaborative filtering and the model-based approach, which are two general QoS prediction approaches.

The System presented in [2] by Hwee Hwa PANG, Xuhua DING and Xiaokui XIAO is a similarity text retrieval framework that bears anonymous security for the query phrases, and thus the client purpose is satisfied, without compromising performance. Their approach is to furnish every client query-phrase with fake terms formerly submitting it to the searching-engine. Beginning from a database of phrase association, they give a technique for choosing fake terms that show comparative specificity spread as the real term, even a point credible to another topic. This likewise, gives a novel retrieval strategy, utilizing homomorphic encryption method that empowers the search-engine to evaluate the encrypted record significance scores concerning just the real search terms, however remain unaware to their differentiation from the decoys.

When user enters any query in an enterprise, that query can disclose the terms in which user is interested and also the confidential or business information. To avoid the revelation of user's true objective behind query terms, it is beneficial to obfuscate the true intension of user. So the system presented in [3] by HweeHwa Pang, Xiaokui Xiao , Jialie Shen provides the approach to outline the terms that are related to user target. They present a TopPriv algorithm to gain the personalized privacy requirement of user by placing automatic generated dummy queries.

Feng Zhang, Victor E. Lee, and Ruoming Jin proposed [4] k-coRating, a novel privacy-protection model to maintain information privacy by substituting some invalid rating with well predicted total Score. They don't just veil the true-ratings, yet additionally improve the information utility, which demonstrates the historical presumption that accuracy and security are two objectives in conflict isn't really right. They demonstrate that the ideal k-coRated mapping is an NP-hard issue and plan a naive however productive calculation to accomplish k-coRating. The significant commitment of this system is demonstrates that the conventional presumption that accuracy and privacy are two objectives in conflict isn't really right. The k-coRating is introduced as an approach to accomplish both higher utility and security. Both the objectives are accomplished by the filling information. The idea is straightforward and also successful.

Yilin Shen and Hongxia Jin proposed the system [5] which is useful for the users when they want to protect their personal data in personalized recommendation. They give the privacy solution for customized recommendation under untrusted server setting in which client's personal information is obfuscated previously flee from their personal device. This system provides greater control of individual on their personal data and mitigates responsibility of service provider on privacy-protection. They provide approach on differential privacy which is the privacy model with slight computation and guaranteed privacy.

Zhifeng Luo, Shuhong Chen, Yutian Li proposed [6] a distribution anonymization which preserves the privacy in recommendation system. This system enables clients to separately anonymize their own particular information without getting to each other's information. In this permutation and multi-pseudonymity are coordinated to anonymize the individual information with the goal that information recognized by the adversary can't be utilized to uncover the private data from the anonymized information. This can be accomplished by the proposed anonymization guide of bipartite graph. Also, this system empowers the anonymized information to preserve the utility of true information for the authorized client while keeping the private data away from the adversary.

III. SYSTEM ARCHITECTURE AND OVERVIEW

Fig. 1 demonstrates the framework structure utilized by the system for the security of users-sensitive choices in a customized suggestion benefit, which comprises an trusted client-sides and untrusted server-side.

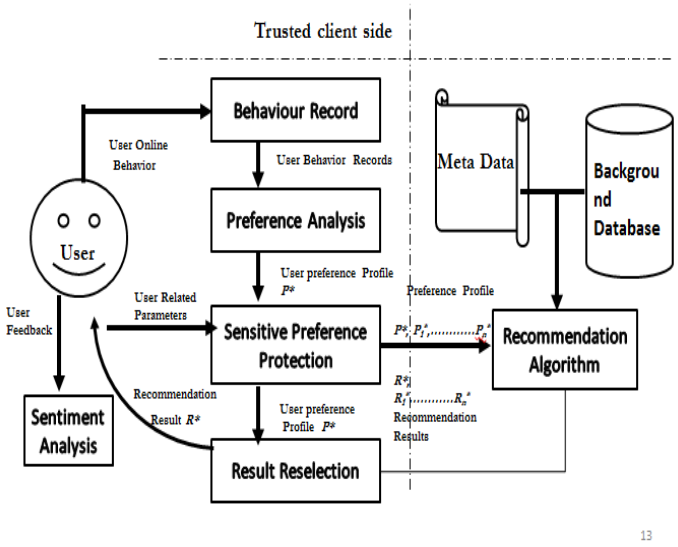

Fig1. Framework structure for preserving privacy in customized recommendation.

Under trusted client side there are following components which play an important role in this system

1). Behaviour Record Component

This is useful to collect the user's online behaviour. It records the user's online activities like searching for any product.

2).Preference Analysis

According to the user behaviour record component this phase analyse the user's preferences. When user likes or purchases any product or item then that product will be the preference of that particular user. Likewise when user stores any product into their cart then also that product can be the preference of that user. So preference analysis is an important part in personalized recommendation.

3).Sensitive preference protection

This phase takes a user related parameter and user preference profile as input and then formulates the dummy preferences based on user's original profile. After this, the dummy profiles are submitted together with original profiles to another side as input to the recommendation algorithm.

4).Result Reselection

Result reselection is important to select the original result which is corresponding to user's original preferences from all recommended result.

5).Sentiment Analysis

Sentiment Analysis is newly introduced component which play an important role to recommend a good quality product to user. This takes a feedback from user and identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the user's attitude towards a particular product is positive or negative.

Actually dummy preferences calculated randomly are so easy to identify, thereby they are unsuccessful to totally cover-up the actual user preferences. So the dummy preferences formulated by sensitive-preference block should meet the requirement of the security of user's personal choices.

A. Algorithm

Following is the algorithm for generating the dummy preferences so as to hide the actual user's preferences which the users don't want to disclose.

Input: (1) F^* user preferences product set, (2) S^+ the user sensitive subjects, (3)Related parameter

Output: $F_1^*, F_2^*, \ldots\ldots\ldots, F_n^*$ a group of dummy product set

begin

From a set of all the subjects S select the subject set with the $1,2,\ldots\ldots k^m$, respectively denoted by $S1,S2,\ldots\ldots S^{km}$ i.e $\forall_g \epsilon S^k \rightarrow$ level(s)= k (k=1,2,... k^m);

From a set of all the user preference subjects S^*, select the subject set with the $1,2,\ldots\ldots k^m$, respectively denoted by $S1,S2,\ldots\ldots S^{km}$;

foreach $S^k \epsilon \{ S1,S2,\ldots\ldots S^{km} \}$ do set $S^k = S^k - S^+$;

Set F=ϕ;

While $\exists_s{}^+ \epsilon S^+ \rightarrow \mu_p .sig(s^+, F^*) < sig(g^+, \{F^*\} U F)$do

Set $F^*_i = \Phi$;

call SearchDummyProducts($S^1, S^*_1, 1, F^*_i$);

set $F = F U \{F^*_i\}$;

return P;

Procedure SearchDummyProduct (Q in, Q^* in, k in, F^*_i in&out)

begin

Select | Q^* | subject from Q randomly to form a dummy subject set $Q^\#$;

Pair the subject in A^* and $A^\#$ randomly

if $k < k^m$ then

foreach $s^* \epsilon Q^*$ do

Let H* be all the subjects in S^*_{k+1} that belong to s^*, and $H^\#$ all the subjects in S^{k+1} which belong to $s^\#$;

call SearchDummyProducts($H^\#$,H*, k+1, F^*_i);

else

foreach $s^* \epsilon Q^*$ do

Let H* be all the products in F^* that belong to s^*, and $H^\#$ be all the products in F that belongs to $s^\#$;

foreach $f \epsilon$ H* do select dummy product f' randomly from $H^\#$ and set score(f') = score(f);

Add all the scored dummy products from $H^\#$ into the dummy product set F^*_i

IV. EXPERIMENTAL DETAILS

There are many components in this system and each component plays an important role to provide the security to personalized recommendation system. fig 2. is a user search module from which user can search for any product and if he/she likes the product they can purchase that item

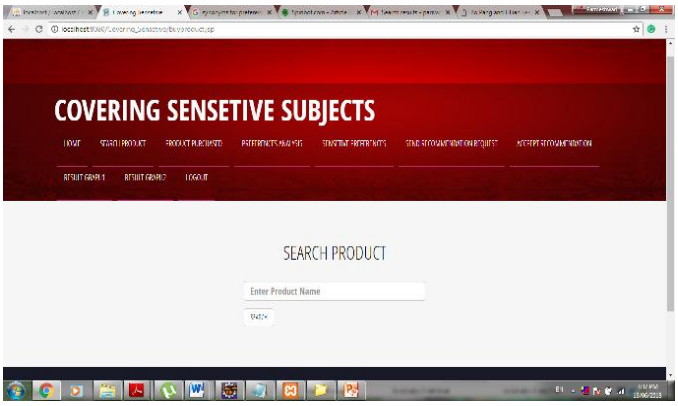

Fig.2. User Search Module

When user search for any product and if he/she gets that product then they can check the details of that product like features of the product, images etc.

Fig 3 demonstrates the preference analysis. When user purchases any product, the preferences of user will record. Preference analysis is so important because this is not only the output of behaviour record but also the input of sensitive preferences.

User can select some preferences as a sensitive preference, which users don't want to disclose. Actual user's choices with dummy choices formulated by algorithm will then transfer to the recommendation algorithm. Recommendation result output by server side is corresponding to actual user choice as well as dummy choice. Again there is result reselection component that will discard the recommendations from dummy preferences.

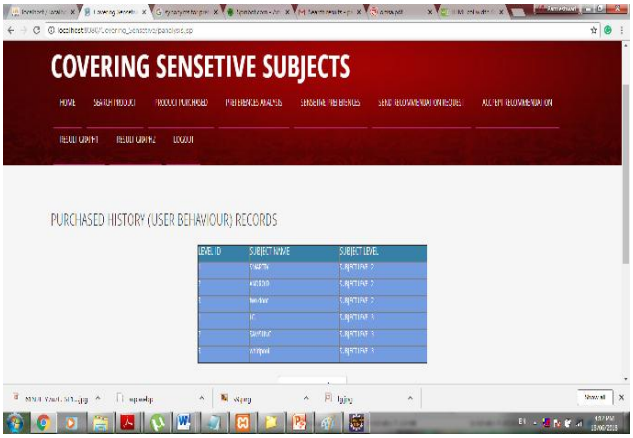

Fig.3. User Behaviour Record

Sentiment analysis is also crucial component in the proposed system which determines the positive and negative attitude of user towards the product. When user gives the review on any product, the sentiment analysis will show that the review as good or bad. So if review is good then it's helpful for other users to purchase that product.

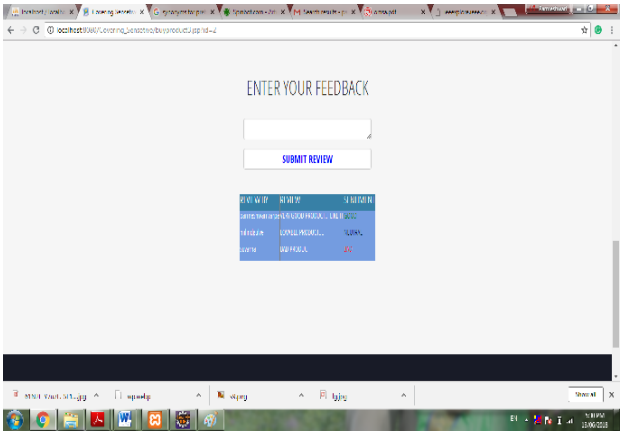

Fig.4. Sentiment Analysis

V. RESULTS ANALYSIS

To provide the security to the user preferences in the recommendation system, it is necessity to produce good quality of dummy preferences. Effectiveness of the approach is depend on conditions that the dummy preference can effectively minimize the significance of sensitive terms and has greatly close feature distribution with user preference set.

First we have to calculate the feature distribution similarity between actual user preferences and dummy preferences. Given an algorithm candidate (A), user preference set F^*, suppose **F** represent a group of dummy preferences formulated for F^*, F_i represent product vector of $F^*_i \in \mathbf{F}$ and S^k_i represent subject-vector with level k=1,2,…. k^m, for F^*_I then equation formulated as

$$\text{ProSim(A)} = \min_{F^*_i \in \mathbf{F}} \{ \text{sim}(F_i, \text{F}) \}$$

$$\text{SubSim(A)} = \min_{F^*_i \in \mathbf{F}} \{ \text{sim}(S^k_i, S^k) \}$$

$$\text{TotalSim} = (\text{ProSim(A)}/ \text{k}^{\text{m}} + 1) + \sum_{K=1}^{k^m} \text{SubSim}_K(\text{A})/ \text{k}^{\text{m}} + 1$$

Maximum value demonstrates the dummy preferences have more similar characteristics as the user preferences, making it hard for attacker to find out the user preferences.

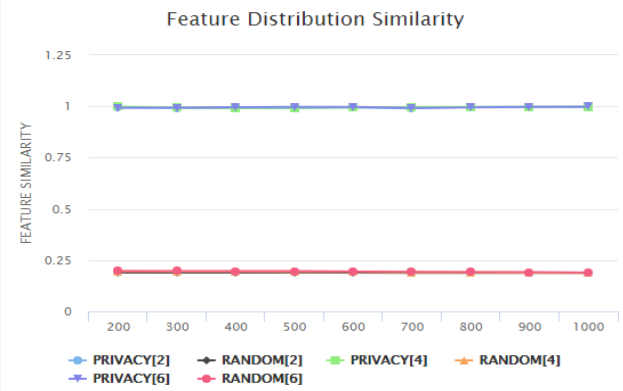

Fig.5. Result for total Fetaure Distribution similarity

Fig.5 shows that privacy approach have more better feature distribution similarity over random approah.

Another evaluation is for significance of sensitive subject is to outline the disclosure degree of sensitive-subject in dummy product set. Significance metric formulated as

$$\text{LevelSignificance}_k(A) = \max_{s^+ \in S^+_k} \frac{\text{sig}(s^+, \{F^*\} \cup \mathbf{F})}{\text{sig}(s^+, F^*)}$$

If it returns the smaller value then it means that the dummy preferences are effectively formulated and they are successful to cover-up the sensitive subject and also making it hard for attacker to identify sensitive subjects. Fig.6 shows that dummy preference set generated by privacy approach can decrease the significance.

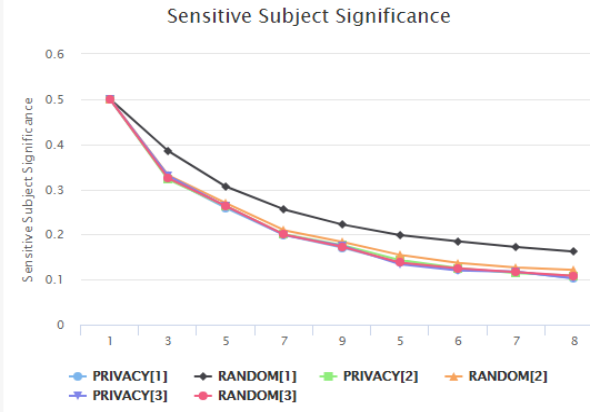

Fig.6. Results for sensitivity subject significance

VI. CONCLUSION

Privacy and Security is an important factor for successful evaluation of personalized recommendation. The proposed work is an approach for securing individual protection for user when utilizing a personalized recommendation benefit, whose fundamental thought is to build a dummy profiles to mask the sensitive subjects contained in a user's-preference profile, and thus to ensure user's individual privacy. As per the result analysis dummy preferences produced by the approach fulfil the requirement of the security of user's actual preferences on untrusted server-side. They decreasing the disclosure degree of user's delicate subject which makes it hard for third party to find out users actual preferences. Proposed approach also provides the review system which is helpful for the user to find out the quality product.

In future work will try to minimize the number of dummy profiles to test the user's privacy. Future work will improve by training very few dummy profiles to analyse user privacy. The system will be tested on different data sets and system will be also updated to improve the security.

ACKNOWLEDGMENT

It gives me an immense pleasure to express my sincere and heartiest gratitude towards my guide Prof. M.A.Wakchaure for his guidance, encouragement, moral support during this work. I am also extremely grateful to my PG Coordinator Prof.S.K.Sonkar for their motivation and support. Finally I would like to thank the Department of Computer Engineering for their moral support.

REFERENCES

[1] Jieming Zhu, Pinjia He, Zibin Zheng, Michael R. Lyu, "*A Privacy-Preserving QoS Prediction Framework for Web Service Recommendation*", 2015 IEEE International Conference on Web Services.

[2] Hwee Hwa PANG, Xuhua DING, Xiaokui XIAO, "*Embellishing Text Search Queries to Protect User Privacy*", Proceedings of the VLDB Endowment: 36th International Conference on Very Large Data Bases: Singapore, 13-17 September 2010.

[3] Hwee Hwa PANG, Xiaokui XIAO, Jialie SHEN, "*Obfuscating the Topical Intention in Enterprise Text Search*", ICDE 2012: IEEE 28th International Conference on Data Engineering, Arlington Virginia, 1-5 April 2012: Proceedings. 1168-1179.

[4] Feng Zhang, Victor E. Lee, and Ruoming Jin, "*k-CoRating: Filling Up Data to Obtain Privacy and Utility*", Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence 2014.

[5] Yilin Shen and Hongxia Jin, "*Privacy-Preserving Personalized Recommendation: An Instance-based Approach via Differential Privacy*", 2014 IEEE International Conference on Data Mining.

[6] Zhifeng Luo, Shuhong Chen , Yutian Li, "*A Distributed Anonymization Scheme for Privacy-preserving Recommendation Systems*", Supported by University Innovation Research and Training Program of Guangdong Province(1056111033) 2013 IEEE.

[7] Zongda Wu, Guiling Li, et al, "*Covering the sensitive subjects to protect personal privacy in personalized recommendation*", IEEE transaction on serviced computing 2016.

[8] Guandong Xu, ZongdaWu, Guiling Li et al. "*Improving contextual advertising matching by using wikipedia thesaurus knowledge*", Knowledge and Information Systems, 2015, 43 (3): 599–631.

[9] Dipasree Pal, Mandar Mitra, Kalyankumar Datta. "*Improving query expansion using WordNet*". Journal of the Association for Information Science and Technology, 2014, 65 (12): 2469–2478

[10] S. Zhang, J. Ford and Fillia Makedon, "*A privacy-preserving collaborative filtering scheme with two-way communication*", Proc. the 7th ACM Conference on Electroinic Commerce, pp. 316-323, 2006.

[11] Liang Hu, Guohang Song, Zhenzhen Xie, and Kuo Zhao, "*Personalized Recommendation Algorithm Based on Preference Features*", Tsinghua science and Technology, Vol. 19, No. 3, 11 pp293-299, June 2014

[12] Yande M, Wakchaure M, Student ME. "*Cross-Site Cold-Start Product Recommendation for Social Media and E-Commerce Websites.*" International Journal of Engineering Science. 2017 Jul;13751.

[13] Khalid O, Khan M U S, Khan S U et al. "*OmniSuggest: A ubiquitous cloud-based context-aware recommendation system for mobile social networks*". IEEE Transactions on Services Computing, 2014, 7 (3):401414.

[14] Varpe P. "*A Preserving Personal Privacy in Personalized Recommendation by protecting the Sensitive Subjects*. "ASIAN JOURNAL FOR CONVERGENCE IN TECHNOLOGY (AJCT)-UGC LISTED. 2018 Apr 15; 4(I).

[15] Shitole MA, Wakchaure MA. "*Patient-Centric and Privacy Preserving Clinical Decision Support System Using Naive Bayesian Classification.*" 2016, pp. 999-1003

[16] Guandong Xu, ZongdaWu, Guiling Li et al. "*Improving contextual advertising matching by using wikipedia thesaurus knowledge*".Knowledge and Information Systems, 2015, 43 (3): 599–631

[17] Wakchaure MM., *Survey on Discrimination Prevention in Data-Mining.*

[18] Mankar A, Patil H, Arage C, Gaikwad M. "*A Survey on Sentiment Computing for the Opinions Based on the Twitter*." International Journal of Scientific Research in computer Science and, Engineering and information Technology, ISSN : 2456-3307, Volume 3 Issue 1, pp.361-364 , 2018

[19] N.Rajganesh, S.Seetha Devi, J. Keerthana, R.Poovizhi, "*A Personalized Job Recommender System Using Hybrid Collaborative Filtering Algorithm*", International Journal of Scientific Research in computer Science and, Engineering and information Technology, ISSN : 2456-3307, Volume 3 Issue 3, pp.192-196 , 2018

Authors Profile

Parmeshwari Varpe is currently pursuing Master Of Engineering in Computer Science from Savitribai Phule Pune University. She has Completed B.E in Computer Engineering. She has published 2 papers in international journal.

.

Prof. M.A. Wakchaure is currently working as Assistant Professor in department of computer Engineering at Amrutvahini College of Engineering Sangamner. He has completed M.Tech in Computer Engineering. Currently he is pursuing Ph.D from Savitribai Phule Pune University. He is having 12 years of teaching experience. He has published 17 papers in international journals and presented 10 papers in international conferences. He fetches various grants from BCUD, Pune.

