IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-5, Issue-10 E-ISSN: 2347-2693

Fine-Grained Knowledge in Agriculture System

Pramod.G. Patil[1*], Sneha.A. Khaire[2]

[1*] Dept. of Computer Engineering, Sandip Institute of Technology and Research Centre, Pune University, Nashik, India
[2] Dept. of Computer Engineering, Sandip Institute of Technology and Research Centre, Pune University, Nashik, India

[*]*Corresponding Author: pgpatil11@gmail.com, Tel.: +91-8407955778*

Available online at: www.ijcseonline.org

Received: 06/Sep/2017, Revised: 17/Sep/2017, Accepted: 08/Oct/2017, Published: 30/Oct/2017

***Abstract*—** For most of the people, web interaction is a very common phase to acquire information. It is possible that in a combined environment, more than one person may try to obtain similar information in one domain. One person may like to solve a problem using an unfamiliar Apache Tomcat which he had studied by another person before. Connecting and then sharing with that persons will be more beneficial to get there learned knowledge. Fine-grained knowledge sharing is proposed for this combined environment. The system is proposed to classify the surfed data into clusters and summarize the details in fine grained details. For any system the efficiency depends upon the surfing. The framework of proposed work includes: (1) Data which is surfed, clustered into tasks. (2)Then task is mined in fine grained output. To get proper result,the search method is applied to the output (mined results).The concept of Data Mining in fine grained knowledge is combined with the information gathering and classification to produce efficient data searching technique in agriculture system.

***Keywords*—**Fine-grain, Cluster, Web-mine.

I. INTRODUCTION

SUPPORT Vector Machines (SVM) shows excellent accuracy to deliver data classification, and accuracy in terms of high performance [1] & [2]. Web Interaction and communication with colleagues are very common routine for knowledge acquitting[4]. Data duplication detection is possible when the data are present in the same real world. Finding the same value in more than one file refers to Data Matching. In data integration, the most essential process is duplication detection. The task of finding entries, which is also called record matching [9] refers to similar entity in two or more files. Problems of duplication detection are solved by performing record matching, that is why the requirements of recognizing the appropriate record matching technique fol- low. Currently used method for duplication de tection in known as supervised methods. For example, several departments in the institution may require the same system software, and the staff of any one department has already have searched about that software. Also in project lab when the new projects come, the developer needs to retrain the previously completed projects background to acquire the back- ground knowledge. When these cases arise, the solution to these can be achieved by restoring the previous data,so that it can save the time and data usages with the benefit of accuracy and minimizing the errors.

The concept of data mining[17] is used in this proposed system,for efficient searching as well as it provides the user an accurate result within data base. In agriculture system there is a need for some mining tools so that the classification of data based on the trained data could be received in minimum time and with great relatively.

In Agriculture system the query matching [5] is proposed to refine the search result as the user trying to retrieve data which are saved in a cluster could be categorise as they could identify them easily for their work [7].

II. RELATED WORK

A. Existing System

Since last decade many developing countries have switched from an industrial age to an information age, and that has drastically changed how we reside, our modes of communication and how we function other operations of our life. Owing to all these changes, now we can access a vast amount of information instantly, through our gadgets. But still these changes are not distributed evenly in our society and this has resulted into poor communication among people, lack of correct information, functions and technology which has affected sustainability in large. Farm management and horticulture is one zone that needs urgent attention to ensure feasibility and accuracy of delivered information [6]. Not only in the developing countries but also the developed countries needs the agriculture field to be enhanced in terms of data availability. Farmers are the untouched element of this field which suffers from poor access to information during the right time of farming cycle to take correct decisions. This proposed system solves this problem by providing the farmer centric information in

such a manner so that any user could access the information based on the agriculture [8].

Drawbacks:

1. Requires deep knowledge of the subject for searching desired information.
2. Increases the overhead of searching non relevant data.
3. Time consuming.
4. Less efficient

B. Associated Research

The agriculture department (DOA) [10] has worked for development and dissemination of better technology in agriculture and focused to acquire sustainable and equitable development of agriculture, through increase in investment in agriculture to attain this goal. For Data classification, Novel Techniques used as with the machine learning. SVM was introduced by Boser, Guyon & Vapnik. Successful applications in many fields like text classification, image recognition[15], face recognition[14], Medical Imaging [16],pattern analysis [13], speaker identification etc.

In 2015 Mudhakar Srivatsa, Ziyu Guan, Huan Sun, Xifeng Yan, Shengqi Yang [23] showed the knowledge sharing by data sharing methods in collaborative Environments. They proposed the two step frame work to refine the generated result. The issue with fine grained data is discussed as the hierarchical tree structure. While using the clustering of relevant data, the data is clustered as the outcome of surfing data by the users. In this case user privacy is an issue. Pradipsinh K. Chavda and JitendraS. Dhobi [24], proposed theA Survey of Model Used for Web Users Browsing Behavior Prediction in Computer Engineering and Intelligent Systems, Vol.6, No.3, 2015, proposed the parallel support vector machine to overcome the scalability issue and also the big data mining problem which is faced by the machine learning algorithms.

Cong Wang , KuiRen , Wenjing Lou and Shucheng Yu proposed the Achieving secure, scalable and fine-grained data access control in cloud computing[25] in which scalability, ease of use and data confidentially is achieved.Iftikhar, N in Integration, Aggregation, and Exchange of Farming Device Data: A high level perspective [26] proposed the solution to integrate, aggregate and exchange of data in effective manner. Data from an agriculture sector is huge and real time

and it is very difficult for processing or analysing. Also speed of generating real time data is high. Algorithms are needed to

process, analyse and make a decision from that BigData to find

semantic analysis in agricultural sector[27]. In 2014, A.Parameshwari, B.Rasina Begum proposed intitled Fine Grained Data Access Control in Cloud Computing [28] proposed the data gathering and analysis of data in the field of medical.

This shows the requirement for a more flexible system which can provide information to farmers such as before plucking the crop or selling their produce to the market. Such system will help farmers immensely. SVM has the advantages which are studied for the system: high prediction accuracy minimizes the errors, works when training examples contain errors, robust.

III. PROBLEM DEFINITION

For a farmer, to take right decision at right time of crop cycle is essential thus the suggested system takes the advantages of Fine grained data mining technique and produces the most relative output. It is obvious that farmers need proper information at right phase of crop cycle to take effective decision. However, a large population is still untouched with development in the field of information technology. The only solution for this problem is to educate farmers about technological development.

A. Objective

The objective of the proposed system is:

1) To justify individuals who are using the desired data in form of knowledge.

2) To provide an active sharing environment.

3) To provide the suitable data to the user with the other users searched data.

4) Time saving and less effort by providing easy access of information.

5) To exploit the generated data by user's behaviour for addressing advisor search.

B. Methodology

The goal of this system is not only to provide the required data but also to classify them in a fine grained manner [18] & [19]. The data classification is done through clusters. The clusters are divided on the basis of time, and are refined with the keyword search. First, the user entered the query in the search panel. The data is processed and extracted. The markov model is developed to fine grain the information and to relate the data with previously searched data[12]. The information retrieval phase in the agricultural system will not be the crucial one. When the users try to access the information about the agriculture management system through the proposed system, they are providing with the Data Mining tool[3], SVM to provide the fined information. Proposed system, based on the basic principle of support vector machine, trains itself with every new query helping the user to get exact information he is searching for. It gives an easy access to the user with previously buffered documents, so that the less time and high accuracy could be achieved.

IV. SYSTEM ARCHITECTURE

To solve a particular problem, the best way will be to apply all available knowledge and spend a respectable amount of time, money and effort in building a required system that will give the right answer[4]. The second best way of doing the same thing is to learn from experience [20] & [21].

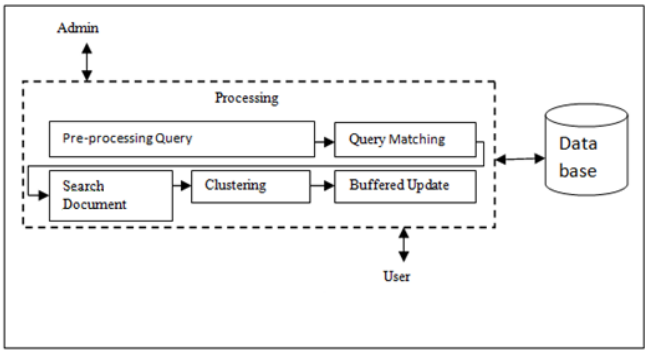

Fig.1. System Architecture

Pre-processing Query: The query generated by the user is processed and the keywords are extracted from the given query. Preprocessing query processes the query for the optimization and analysis the query as set of tokens.

Query Matching: The pre processed query is then took as a input for query matching module, which further matches each of the tokens with the keyword from the documents.

Search Document: The matched query is then searched in the documents. If the document is present in the buffer, the documents will be provided from there. If not the document will be searched in the main database. The result will be displayed as sequence of satisfied documents.

Clustering: The clustering is an important phase in the system. As the documents are clustered with their type of information also with the property of time, i.e. the documents
are clusters on the basis of their month of uploading. The documents will be clustered with their area like fertilizers; crop, soil etc. Also will be displayed as the packets of documents as month wise clustering.

Buffered Update: The buffering of recent searched document should be updated every time a user searches the new document. After the searching of particular document, the buffer is updated with its keywords so that the next search could relate with the buffered items. After completing the intensive study of data following problem are identified:-

1. Various website provides Horticulture information but the less accuracy and time consumption is the measure issue.
2. Existing system does not provide the information in different prospective.
3. While using any service provider to extract some specific knowledge, it generates unwanted links regardless of the searched query.

The main components of this architecture are: input, processing, output.
Input: The Input for the system will be the query asked by the user to get the information. Here the query is treated as an instruction.

Processing: The Processing part consist of many phases.

1. The instruction is selected for the entry
2. The keywords are then searched in the table
3. The table consist of the buffer register,checks for the relevant searches using SVM
4. If the relevant search is available in the buffer, the output is displayed directly to the user.
5. If no relevant data is available in the buffer, the query extraction is performed on the database.
6. The output is displayed.

V. PROBLEM FORMULATION

The input for the system is the query generated by the user, which processed and deliver the result as refine data. For processing, the data value is extracted, selected and then check for the buffer for data availability.

A. Algorithm for Support Vector Machine

Steps:

1. Start
2. Train sample documents related to each of the classes
3. Identify keywords from each of the Data.
4. Calculate term weight of each keyword in Data
5. Store stemmed keywords and term weight (learned function) in Data Base.
6. Calculate avg weight of document using term weights
7. Give above parameters as input to SVM training module
8. Generate training model for each of the classes(n)
9. For each category x
10. Identify keywords entered by the user (search)
11. Retrieve weight from database of selected category x (can check at file level)
12. Calculate avg weight (or multiple avg for loop) of keywords using term weights for all keywords entered by user
13. Analyse predicted values for each of the classes

14. Display list of classes associated with current document
15. Stop

B. Mathematical Model

The working of system with the data flow is shown in terms of mathematical model.

1. Processing

Let,

I be the set of input containing text and its properties. $I = I_1 , I_2 , I_3 .....I_N$

O be the set of output containing processed text. $O = O_1 , O_2 , O_3 .....O_N$

F = set of functions $F = F_1 , F_2 , F_3 , F_4$ .....

F 1 = Selected entry
F 2 = Reprocessing
F 3 = Data Extraction
F 4 = Data Selection
F 5 = Training Data
F 6 =Testing Data

2. Classification

Classification of data as the information should belong to a cluster, the data classification is needed.

Let,

M be the Margin width
wx+b=0, Seperator
wx+b=+1, Predict class =+1 zone,
wx+b=-1, Predict class = -1 zone,
So, $w.x^{+} + b = +1$
$w.x^{-} + b = -1$
$w.(x^{+} - x^{-}) = 2$
$M = \frac{(x + - x -).w}{|w|} = 2/|w|$

3. Problem Description

Let S be a Data Mining tool; Such that
S = {N, R, J, L, H, C, S, E, K | ∅ s{ where
N represents normal user;
N = { | ∅ n } and
R represents registration of user;
R = {r 0 , r 1 .....r n | ∅ r } and
J represents profile created by user;
J = {j 0 , j 1 , ...j n | ∅ j } and
L represents request;
L = {l 0 , l 1 | ∅ l } and
H represents get all data;
H = {h 0 , h 1 , ...h n | ∅ h } and
C represents mapping;
C = {c 0 , c 1 , ...c n | ∅ c } and
S represents extraction;
S = {s 0 , s 1 , ...s n | ∅ s } and
E represents selection;
E = {e 0 , e 1 , ...e n | ∅ e } and
Y represents keyword match;
Y = {y 0 , y 1 , ...y n | ∅ y .
Let f u be a rule of N into S such that each user sends request to the Server.

$f_n (n_0 , n_1 , ..., n_n) \mapsto \{s_0\} S$
$f_j (j_0 , j_1 , ..., j_n) \mapsto \{s_0\} S$
$f_h (h_0 , h_1 , ..., h_n) \mapsto \{s_0\} S$
$f_c (c_0 , c_1 , ..., c_n) \mapsto \{s_0\} S$
$f_e (e_0 , e_1 , ..., e_n) \mapsto \{s_0\} S$
$f_l (l_0 , l_1 , ..., l_n) \mapsto \{s_0\} S$

VI. CONCLUSION

The method SVM provides an easy way for the user of agriculture management system to retrieve their data in a fine grained manner with advisor search by exploring the past relevant results. In agriculture system, all the users of this system are not technically literate. The goal of the user will be to get the refine result by the proposed system in reflection of peoples interactions.SVM will provide accurate and robust classified results. The proposed work could be enhanced further by applying this concept to image data, which will be the add-ons for the farmers to check the information in graphical form. The limitations of hierarchical tree structure of the algorithm can be resolve by using hybrid algorithms.

REFERENCES

[1] Boser, B. E., I. Guyon, and V. Vapnik (1992), "*A training algorithm for optimal margin classifiers*", . In Proceedings of the Fifth Annual Workshop on Computational Learning

[2] V. Vapnik, *The Nature of Statistical Learning Theory*, NY: Springer-Verlag. 1995.

[3] Meijuan Gao, Jingwen Tian, Shiru Zhou, "*Research of Web Classification Mining Based on Classify Support Vector Machine* ", SECS International Colloquium on Computing, Communication, Control, and Management -2009.

[4] Wang, P., Hu, J., Zeng, H. J., & Chen, Z. (2009). "*Using Wikipedia knowledge to improve text classification*", Knowledge and Information Systems, 9-13.

[5] Wei Wu & Zhengdong Lu Hang Li ,"Learning Bilinear Model for Matching Queries and Documents",

[6] Krishi-Mitra: "*Expert System for Farmers*" ,IJCSMC, Vol. 4, Issue. 4, April 2015, pg.893 899

[7] Cambazoglu B. B., Aykanat C.:" *Performance of query processing implementations in ranking-based text retrieval systems using inverted indices*". International Journal of Food and Agricultural Economics ISSN 2147-8988 Vol. 1 No.1 pp. 63-74.

[8] *Federal Committee on Statistical Methodology (1980),* Statistical Policy Working Paper 5: Report on Exact and Statistical Matching Techniques Washington, DC: Office Federal Statistical Policy and Standards, U.S. Department of Commerce. "Organizational History of the Department of Agriculture & Cooperation. Retrieved" 5 July 2012.

[9] Z. Pawlak, "*Rough sets and intelligent data analysis",* Information Sciences, 147 (2002) 1 12.

[10] K.I. Kim, K. Jung, S.H. Park, H.J. Kim, "*Support vector machines for texture classification*", IEEE Transactions on Pattern Analysis and Machine Intelligenc 24(11) (2002) 15421550.

[11] Latha Parthiban, "*A Novel Face Recognition Algorithm with Support Vector Machine Classifier*", International Journal of Mathemetics and Scientific Computing", Vol. 1,no. 1,2011.

[12] R.Muralidharan , and Dr.C.Chandrasekar, *"Object Recognition Using Support Vector Machine Augmented by RST Invariants*"IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, September 2011.

[13] Mohd. Aquib Ansari, Diksha Kurchaniya, Manish Dixit, Punit Kumar Johari, "*An Effective Approach to an Image Retrieval using SVM Classifier*", International Journal of Computer Sciences and Engineering, Vol.5, Issue.6, pp.62-72, 2017.

[14] R. Baker and K. Yacef, "*The State of Educational Data Mining in 2009: A Review and Future Visions*', J. Educational Data Mining, vol. 1, no. 1, pp. 3-17, 2009.

[15] J. Kaur, S.S. Sehra, S.K. Sehra, "*A Systematic Literature Review of Sentiment Analysis Techniques*", International Journal of Computer Sciences and Engineering, Vol.5, Issue.4, pp.22-28, 2017.

[16] Uma Aggarwal, Gaurav Aggarwal , "*Sentiment Analysis on Demonetization using SVM*", International Journal of Computer Sciences and Engineering, Vol.5, Issue.6, pp.183-187, 2017.

[17] Binder, J., Murphy, K., & Russell, S. (1997b). *Space-efficient inference in dynamic probabilistic networks.* In Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97) Nagoya, Japan. Morgan Kauf-mann.

[18] A. M. Turing, *Intelligent Machinery*, in Machine Intelligence, B. Meltzer and D. Michie, Eds. Edinburgh: Ed-inburgh University Press, 1969, vol. 5, National Physical Laboratory Report (1948).

[19] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, *A Practical Guide to Support Vector Classification*.

[20] Ziyu Guan, Shengqi Yang, Huan Sun, Mudhakar Srivatsa,Xifeng Yan,*Fine-Grained Knowledge Sharing in Collaborative Environments* , IEEE Transactions on Knowledge and Data Engineering,2015.

[21] Pradipsinh K. Chavda and Jitendra S.Dhobi , *proposed the A Survey of Model Used for Web Users Browsing Behavior Prediction in Computer Engineering and Intelligent Systems*, Vol.6, No.3, 2015.

[22] Cong Wang , Kui Ren , Wenjing Lou and Shucheng Yu *proposed the Achieving secure, scalable and fine-grained data access control in cloud computin*.,INFOCOM, 2010 preceedings IEEE.

[23] Iftikhar, N, Integration, *Aggregation, and Exchange of Farming Device Data: A high level perspective*, Application of digital information and web Technologies, 2009. ICADIWT09.

[24] Rakesh. S.Shirsath, Vaibhav A.Desale, Amol. D.Potgantwar, "*Big Data Analytical Architecture for Real-Time Applications*", International Journal of Scientific Research in Network Security and Communication, Vol.5, Issue.4, pp.1-8, 2017.

[25] Pramod.G. Patil, Sneha.A. Khaire, "*Fine-Grained Knowledge In Agriculture System*", International Journal of Computer Sciences and Engineering, Vol.5, Issue.10, pp.287-292, 2017.

Authors Profile

Mr. P. G. Patil completed Bachelor degree of Computer Engg from North Maharashtra University in the year 2010 and Master's from RGVP Bhopal University, in year 2016. Currently working as Assistant Professor in Department of Computer, University of Pune. Main research work focuses on Big Data Analytics, Network Security, Data Mining. Main research work focuses on Cryptography Algorithms, Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining, IoT, Wireless computing and Computational Intelligence based education. Have 5 years of teaching experience. Also published books on Computer Networks and Computer Graphics.

Miss. S. A. Khaire completed Bachelor degree of Computer Engg from Pune Universityin the year 2014 and Master's from the same University, in year 2017. Currently working as Assistant Professor in Department of Computer, University of Pune. Main research work focuses on Big Data Analytics, Network Security, Data Mining.

