International Journal of scientific research and management
(IJSRM) ||Volume||2||Issue||5||Pages|| 825-829 ||2013||
Website: www.ijsrm.in ISSN (e): 2321-3418

Data incremental approach for Dynamic environment using Rough set theory

Ashok Kumar.S[1], Gopinath.S[2]

[1]Gnanamani College of Technology, Department of Computer Science and Engineering,
Namakkal 637018, India
write2ashokkumar@gmail.com

[2]Gnanamani College of Technology, Department of Computer Science and Engineering,
Namakkal 637018, India

Abstract. Approximations of a concept by a variable precision rough-set model (VPRS) usually vary under a dynamic information system environment. It is thus effective to carry out incremental updating approximations by utilizing previous data structures. This paper focuses on a new incremental method for updating approximations of VPRS while objects in the information system dynamically alter. It discusses properties of information granulation and approximations under the dynamic environment while objects in the universe evolve over time. The variation of an attribute's domain is also considered to perform incremental updating for approximations under VPRS. Finally, an extensive experimental evaluation validates the efficiency of the proposed method for dynamic maintenance of VPRS approximations. The variation of an attribute's domain is also considered to perform incremental updating for approximations under the system walls. Finally, an extensive experimental evaluation validates the efficiency of the proposed method for dynamic maintenance of VPRS approximations. This is achieved through a flexible rule-based system, that allows users to customize the filtering criteria to be applied to their walls, and a Machine Learning based soft classifier automatically labeling messages in support of VPRS filtering.

Keywords: Decision-Theoretic Rough Set model, Machine Learning, Overall Accuracy, Online Social Networks, Prediction by partial matching.

1. INTRODUCTION

The first activity in the phase is studying the existing system and other is to understand the requirements and domain of the new system. Both the activities are equally important, but the first activity serves as a basis of giving the functional specifications and then successful design of the proposed system. Understanding the properties and requirements of a new system is more difficult and requires creative thinking and understanding of existing running system is also difficult, improper understanding of present system can lead diversion from solution.

The huge and dynamic character of these data creates the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant within the data. The auditor verifies the integrity of the data file and the server's possession. This scheme only works for the auditor to maintain state, and suffers from bounded usage, which potentially brings in online burden to users when the repeated values are used up.

In other related work, thoroughly study set of requirements which ought to be satisfied for a remote data possession checking protocol to be of practical use. Their proposed protocol supports unlimited times of file integrity verifications and allows preset tradeoff between the protocol running time and the local storage burden at the user. Most importantly, in that they never reveal the data contents to the auditor. Our solution removes the burden of verification replicated data, alleviates both the customer's and storage service's fear of data leakage, and provides a method for independent arbitration of data retention contracts.

2. RELATED WORK

A Decision –Theoretic Rough set model to require the ML (Machine Learning) paradigm according to which a classifier is automatically induced by learning from a set of preclassified examples. A remarkable variety of related work has recently appeared which differ for the adopted features extraction methods. Which differ for the adopted feature extraction methods, model learning, and collection of procedure maps text. Several experiments prove that Bag of words approached yield good performance and prevail in general over more sophisticated text representation that may have superior semantics but lower statically quality. In particular the derived well known overall accuracy (OA) index capturing the simple percent agreement between truth and classification is complemented to the system. A content based approaches can posted on environments poses additional challenges given the short length of these messages other than wide range of topics that can be included.

The Proposals exploiting classification mechanisms for personalizing access in VPRS classification method has been proposed to categorize short text messages in order to avoid overwhelming users of micro blogging services by raw data. The user can then view only certain types of tweets based on his/her interests. The only social networking service we are aware of providing filtering abilities to its users in the subscriber. They are instrumental to provide an active support in complex and sophisticated tasks involved in VPRS management, such as for instance access control or information filtering. Information filtering has been greatly explored for what concerns textual documents and, more recently, web content. Finally our policy languages has some relationship with the policy frameworks that have been so far proposed to support the specification and enforcement of policies expressed in terms of constraints on the machine understand able resource description provided by semantic web languages.

We address short text categorization as a hierarchical two level classification process. The first level classifier performs a binary hard categorization that labels messages as neutral and non-natural. The first level filtering task facilitates the subsequent second level task in which a finger grained classification is performed. The second level classifier performs a soft partition of Non-neutral messages assigning a given message a gradual membership to each of the non-neutral classes. The difficulty has been well addressed by DTRS since the threshold can be obtained by the loss function in DTRS. When the information system evolves with time, there are a few works on knowledge updating in VPRS. Wang et al. discussed the relation between a new record and the existing equivalence classes on the condition attributes and effects on the rule set caused by the new record in VPRS. A growing number of online services, such as Google, Yahoo!, and Amazon, are starting to charge users for their storage. Customers often use these services to store valuable data such as email, family photos and videos, and disk backups. Today, a customer must entirely trust such external services to maintain the integrity of hosted data and return it intact. Unfortunately, no service is infallible. To make storage services accountable for data loss, we present protocols that allow a third-party auditor to periodically verify the data stored by a service and assist in returning the data intact to the customer. Most importantly, our protocols are privacy-preserving, in that they never reveal the data contents to the auditor. Our solution removes the burden of verification from the customer, alleviates both the customer's and storage service's fear of data leakage, and provides a method for independent arbitration of data retention contracts.

We develop a rigorous, unified framework based on ordinary differential equations (ODEs) to study epidemic routing and its variations. These rough based can be derived as limits of Markovian models under a natural scaling as the number of nodes increases. While an analytical study of Markovian models is quite complex and numerical solution impractical for large networks, the corresponding VPRS models yield closed-form expressions for several performance metrics of interest, and a numerical solution complexity that does not increase with the number of nodes. Using this VPRS approach, we investigate how resources such as buffer space and the number of copies made for a packet can be traded for faster delivery, illustrating the differences among various

forwarding and recovery schemes considered. We perform model validations through simulation studies. Finally we consider the effect of buffer management by complementing the forwarding models with Markovian and fluid buffer models.

The lack of benchmarks for VPRS short text classification makes problematic the development of a reliable comparative analysis. However an indirect comparison of our method can be done with work that shows similarities or complementary aspects with our solution. The Results of analysis conducted in the representative power of the three type of features tailed in general with our VPRS server walls. Which differ for the adopted feature extraction methods, model learning, and collection of procedure maps text. Several experiments prove that approached yield good performance and prevail in general over more sophisticated text representation that may have superior semantics but lower statically quality. A short text classification using a statistical model named prediction by partial matching (PPM).

3. SYSTEM DESIGN

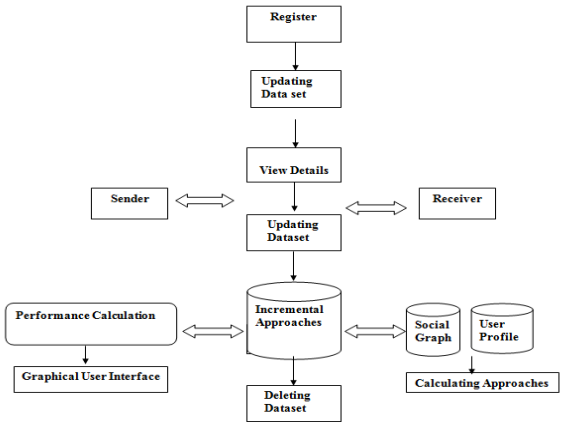
The Variable Precision Rough Set model (VPRS) proposed by Ziarko gives a classification strategy in which the error rate is less than a given threshold . The VPRS has been solving noise data problems with great success in many applications although a difficulty with VPRSis the estimation of the threshold. The difficulty has been well addressed by DTRS since the threshold can be obtained by the loss function in DTRS.

When the information system evolves with time, there are a few works on knowledge updating in VPRS. Wang et al. discussed the relation between a new record and the existing equivalence classes on the condition attributes and effects on the rule set caused by the new record in VPRS. They further proposed an incremental rule acquisition algorithm based on VPRS. In, Liu et al. constructed an accuracy matrix and a coverage matrix under VPRS. The designed system meets the end user requirement with economical way at minimal cost within the affordable price by encouraging more of proposed system. Economic feasibility is concerned with comparing the development cost with the income/benefit derived from the developed system. In this we need to derive how this project will help the management to take effective decisions. As far as the learning model is concerned, we confirm in the current paper the use of neural learning which is today recognized as one of the most efficient solutions in text classification. In particular, we base the overall short text classification. Then they proposed an incremental approach to update the accuracy matrix and coverage matrix to obtain interesting knowledge w.r.t. the immigration or emigration of objects.

*Login and Registration: U*ser can register their details like name, password, gender, age, and then. Here the user can make friends by accept friend request or send friend request. They can share their status by messages also share videos with friends and get comments from them.

Variable Precision Rough Set model: The Variable Precision Rough Set model (VPRS) proposed by Ziarko gives a classification strategy in which the error rate is less than a given threshold. The VPRS has been solving noise data problems with great success in many applications although a difficulty with VPRSis the estimation of the threshold. The difficulty has been well addressed by DTRS since the threshold can be obtainedby the loss function in DTRS.

When the information system evolves with time, there are a few works on knowledge updating in VPRS. Wang et al. discussed the relation between a new record and the existing equivalence classes on the condition attributes and effects on the rule set caused by the new record in VPRS. They further proposed an incremental rule acquisition algorithm based on VPRS.



3.1 Incremental Approach Architecture

Knowledge discovery: The knowledge levels are formed when different attribute sets are utilized to partition the universe. The root of the knowledge hierarchical tree is composed of all elements of the universe U. Sub node is composed of equivalence classes, which are formed via partitions by the attribute set Ai. Each equivalence class is divided into smaller equivalence classes by adding more attributes. In such a way it is, for instance, possible to define rules applying only to young creators or to creators with a given religious/political view. Given the social network scenario, creators may also be identified by exploiting information on their social graph. This implies to state conditions on type, depth and trust values of the relationship(s) creators should be involved in order to apply them the specified rules.

Granular Computing: Granular Computing (GrC) based on Zadeh's "information granularity" describes and processes uncertain, vague, incomplete, and mass information. By using granules and relationships between granules, the focus on certain granularity, and transfer freely between different granularities are realized. Problems can be solved at different granularities' levels of GrC. Yao et al. described basic issues and methods of GrC.The frameworks, models, methodologies, and techniques of GrC were studied. By now GrC has been successfully used in knowledge discovery. Rough Set Theory (RST), one of the leading special cases of GrC approaches, is a formal mathematical theory that models knowledge about the domain of interest in terms of a collection of equivalence relations. Each equivalent class of RST may be viewed as a granule since the elements in the same equivalence class are indistinguishable.For any subset of the universe is described approximately by the equivalence classes. The granulation structure is thus induced by the equivalence relation in RST.

Incremental Updating: Incremental updating is an effective method to maintain knowledge dynamically and it has been used to data analysis in the real-time applications,e.g., steam data, interactive application, multiple learning agents, and the applications with a limited memory or Computation ability. For example, Kang etal. Proposed a novel algorithm for incremental and general evaluation of continuous reverse nearest neighbor queries. Altiparmak et al. proposed an algorithm to update synopsis over a sliding window of most recent entries dynamically in constant time. There has been much research on dealing with incremental updating in RST. When the object set remains unchanged and a single attribute is added into or deleted from the system, An et al. proposed an algorithm for incremental updating approximations and a method for extracting rules

Information System: One of our future study is to consider the objects to be added/deleted as a whole and develop algorithms for updating approximations under VPRS. Another future study will focus on an extension of the algorithms to DTRS, incomplete information systems and other extended rough sets models. The variation of attributes in the information system may further be taken into consideration in terms of incremental updating knowledge.

CONCLUSION

The incremental methods for updating approximations under VPRS when the information system is updated by inserting or deleting an object. The methods are the first effort to efficiently update approximations. After having discussed the principle of updating equivalence classes, we proposed the incremental methods for updating approximations under VPRS in terms of inserting or deleting an object.

By considering the change of a concept and the attribute domain we discussed the alteration of knowledge granulation the variation of data sets. At last, we carried out extensive experiments to verify the effectiveness of the proposed algorithms. Also concluding that the proposed work can focus network message delivered based on incremental approaches walls automatic removal of unnecessary messages from buffer overflow in filtered walls. The solutions investigated in this paper are an extension of those adopted in a previous works from which inherit the learning model and the elicitation procedure for generating pre-classified data. One of our future study is to consider the objects to be added/deleted as a whole and develop algorithms for

updating approximations under VPRS. Another future study will focus on an extension of the algorithms to DTRS, incomplete information systems and other extended rough sets models. The variation of attributes in the information system may further be taken into consideration in terms of incremental updating knowledge .Experimental results show that the algorithms are effective to maintain knowledge when the object set in the information system varies over time.

REFERENCES

[1] A. Adomavicius and G. Tuzhilin, " A Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.

[2] M. Chau and H. Chen, "A Machine Learning Approach to Incremental approaches Using Content and Structure Analysis," Decision Support Systems, vol. 44, no. 2, pp. 482-494, 2008.

[3] R.J. Mooney and L. Roy, "A Rough Based system can perform the incremental approaches," Proc. Fifth ACM Conf. Digital Libraries, pp. 195-204, 2000.

[4] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1-47, 2002.

[5] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari, " Avariable precision rough-set model (VPRS" Proc.ECML/PKDD Workshop Privacy and Security Issues in Data Mining and Machine Learning (PSDML '10), 2010.

[6] N.J. Belkin and W.B. Croft, "Information detecting and Information Retrieval: Two Sides of the Same Coin?" Comm. ACM, vol. 35, no. 12, pp. 29-38, 1992.

[7] P.J. Denning, "Electronic Junk," Comm. ACM, vol. 25, no. 3, pp. 163-165, 1982.

[8] P.W. Foltz and S.T. Dumais, "Personalized Information Delivery: An Analysis of Information Filtering Methods," Comm. ACM, vol. 35, no. 12, pp. 51-60, 1992.

[9] P.S. Jacobs and L.F. Rau, "Scisor: Extracting Information from On- Line News," Comm. ACM, vol. 33, no. 11, pp. 88-97, 1990.

[10] S. Pollock, " Scisor Rule-Based incremental approaches," ACM Trans. Office Information Systems, vol. 6, no. 3, pp. 232-254, 1988.

AUTHOR PROFILE

Ashok Kumar.S received the B.E degree in Computer science and Engineering from P.G.P College of Engineering and Technology in 2005. He is pursuing towards the M.E degree in Computer Science and Engineering from Gnanamani College of Technology, Affiliated to Anna University, Chennai since September 2012. His research area is Database Management Systems and Data Warehousing.

Gopinathi.S received the B.E degree in computer Science and Engineering from Selvam College of Technology, in 2010, and M.E degree in Computer Science and Engineering from Jayaram College of Engineering and Technology, in 2012, now working as Assistant Professor in Gnanamani College of Technology. His research area includes Database Management Systems.