International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 05 | May -2020 ISSN: 2582-3930

Recommendation System for Attribute Access Control of Unstructured Text

Prof. A. V. Deorankar[1] and Kavita B. Gawande[2]

[1] *Department of Computer Science and Engineering, Government College of Engineering Amravati*
[2] *Department of Computer Science and Engineering, Government College of Engineering Amravati*

Abstract - Attribute based access control (ABAC) is used to provide the privacy and security to data. Now a days many platforms are used access control mechanisms for their application. For ABAC model, attributes are the basis for controlling access to data. The existing attribute extraction methods that are based on manual management, which are time consuming and have a high cost, variable accuracy, and poor scalability when dealing with massive unstructured text. The method are required for carefully calculates the characteristics of attributes themselves, the relationships between attributes, and the relationship between attributes and resources. This paper proposes a feature generation method for text resource attributes. It can fully and accurately characterize the attributes. By using support vector machine attribute are classified which is helpful for access the proper data. This avoids the need for security experts to manually label the attributes of massive resources and facilitates the automatic and intelligent mining of ABAC resource attributes. We also provide the Recommender systems, whose aim at providing recommendations for services that are targeted to specific users.

Keywords:Attribute-based access control (ABAC), attribute mining, machine learning, unstructured data, recommendation system.

1. INTRODUCTION

Big data technology has received more and more attention. While big data can provide convenience for people, for example, by suggesting things they may like, it also brings huge security and privacy risks. For providing big data security, Access control technology is one of the best technologies [1]. Access controls prevent the unauthorized use of data resources by managing user permissions.

In the 1990s, the role-based access control model has beenstudied continuously, since 1996, after Sandhu et al. proposedthe RBAC model[2], scholars proposed a series of improvedmodels based on this model, which makes the management usershave more convenient permissions, reduce the complexity ofsystem rights management. More than 20 years later, the role basedaccess control model still occupies more than half of theaccess control system.With the continuous development of big data and cloudcomputing, early role-based access control systems arebecoming larger and larger, and the amount of data increases,and the maintenance cost of access control systems increasesrapidly. When the user puts forward new permissionrequirements, the original role-based access control system isdifficult to maintain, and the simple role access control, itsflexibility and control fine grain can't meet the actual needs.

Role-based access control (RBAC) is very widely used but has notable limitations, prompting a shift towards attribute based access control (ABAC), which allows policies to be written in a more flexible and higher-level way. However, the cost of developing an ABAC policy can be a significant obstacle to migration from RBAC to ABAC. Policy mining algorithms can significantly reduce this cost, by partially automating the construction of an ABAC policy from an RBAC policy with accompanying data about attributes of users and resources.

Attribute-Based Access Control (ABAC) uses the attributes of the subject, resource, operation and environment as the basic elements of access control. It can flexibly utilize the set of attributes owned by the subject and or the resource to determine whether to grant access or not. Attributes can strongly express semantics. ABAC can better separate policy management from the access control decision and is compatible with traditional mechanisms such as discretionary, mandatory and role based access control [3],[4]. It is suitable for solving fine-grained access control and large-scale dynamic authorization in big data computing environments.

2. RELATED WORK

In the field of attribute mining there are many papers on focus role mining [5],[6], but paper on resources attribute mining are minimum. A new challenge is to develop new techniques and systems to extensively exploit the large amount of data – the big data. Many information management architectures have been developed towards this goal. Access control, as first introduced to RDBMS in System R [7], is an integral component in legacy information systems.

Content-Based Access Control (CBAC) model is useful for sharing content-centric information of big data. It is basically for content-centric information shearing. CBAC makes decision on the basis of similarity between the content of data and requester's credential [8].

RBAC technologies are used in most of the data base system and enterprise information system. Security mechanism is main part in access control technologies. Role access control is simple. Because of its flexibility and control granularity sometimes it can't get actual access control requirements [9]. Role-based Access Control is mostly use, but there is some limitation. That's why design an algorithm to get ABAC policy from RBAC policy. It gives the overall definition, feature, and attributes data from RBAC [10].

There is problem to deal with noisy and unbalanced training data for supervised key phrase extraction. We get data from many sources with many annotators per document. Simple but useful solution have to provide with unbalanced data without loss of status of unlabeled candidate phrases. Key phases annotation is mostly useful [11].

Now a days, in social media there is rapid growth of data. This data source is useful for the research of big data analysis. In that data text mining is use for applications. Support vector machine (SVM) is use for classification in text mining, also text processing, feature selection methods are used. There is number attribute on dataset, to reduce the problem of attribute classification and feature selection used the Information Gain [12]. Traditional classification algorithms have limitation when it perform with unbalanced data sets. In this situation, we have to focus to modify the SVM algorithm which modification will help to solve the problem of class imbalance. We can add new parameters in SVM [13].

Recommender systems in academia help researchers and scientists overcome information overload. However, only a few operators of academic services, such as digital libraries and reference managers, offer recommender systems to their users. Mr. DLib's recommendations-as-a-service, which allows third parties to easily integrate a recommender system into their products [14].

3. METHODOLOGY

The proposed framework as shown in fig. 1 processes review dataset and summarize it for better reading. The system consists of two section i.e. admin section and user section. Framework consists of different stages to carry out the internal working of system. Attribute mining mechanism use for attributes which is for access control. For better classification machine learning algorithm is used.

3.1 User Section

The user can register and login in the system. The user can search the interested data according to their attributes. It can also get the recommended data according to user's interest.

3.2 Admin Section

The admin can see the list of all registered user. It can also block or allow any user. It can handle all the data and processes which is proceed for attribute based access control.

Access control attribute mining framework consist of following process:

(1)Preprocessing. It is includes removal of redundant and invalid words, part-of-speech tagging. The dataset is cleansed so that further processing can be done effectively. All the white spaces and special characters are removed. Also, the reviews are split into each individual sentences.

(2) Generate a candidate attribute set. Text, stop word and words of particular part-of-speech are removed on the bases of attribute semantic association analysis. The candidate attribute set is extracted using TF-IDF feature. The TF-IDF value evaluates the importance of an attribute, by calculating the frequency with which the attribute appears in the current document and the frequency with which the attribute appears in different documents.

(3) Feature selection. Convert the word into vector. Selected the required attribute according to features of attribute, relationship of attributes and relationship of resources and attributes.

(4) Attribute mining. For this used SVM classification method to classify the attribute. SVM is performed on candidate word vectors to classify them into pre-determined categories. Categories determined from classification are the set of their respective articles.

(5) Post-processing. The interest of user are matched with the categories of articles and those articles are shown to user. That means user can get the article only which have permission. Unauthorized user cannot get articles.

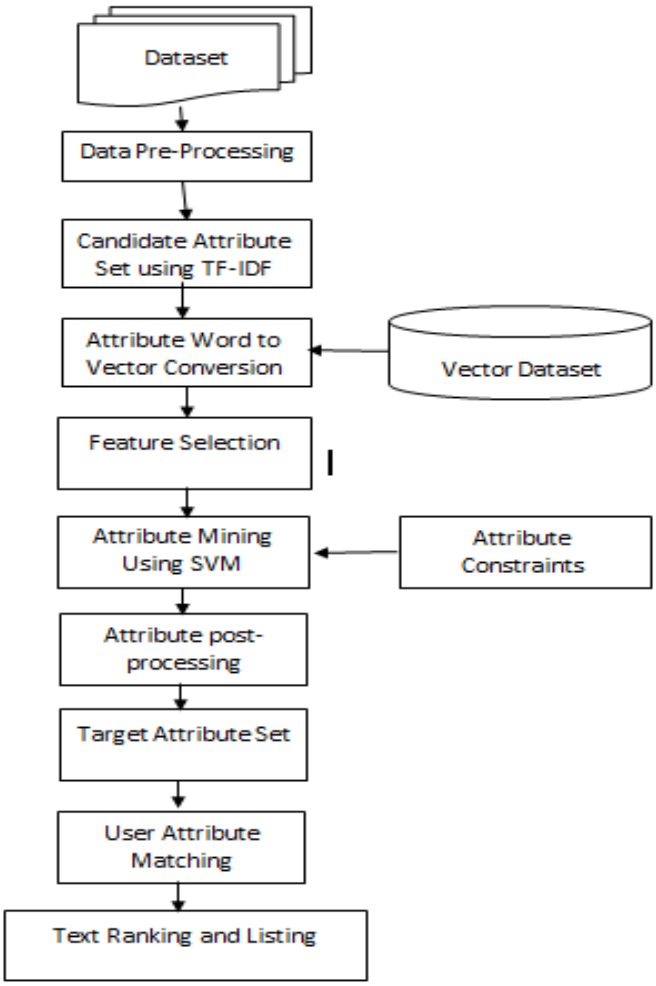

Fig -1: System architecture

4. RESULT

This section includes our results. The result shows the overall performance of the system for classification of attribute. Here we use the attribute finance, sports, technology, nature and

Miscellaneous. The result of the system is measured by the parameters Recall, Precision, F-Measure and Accuracy.

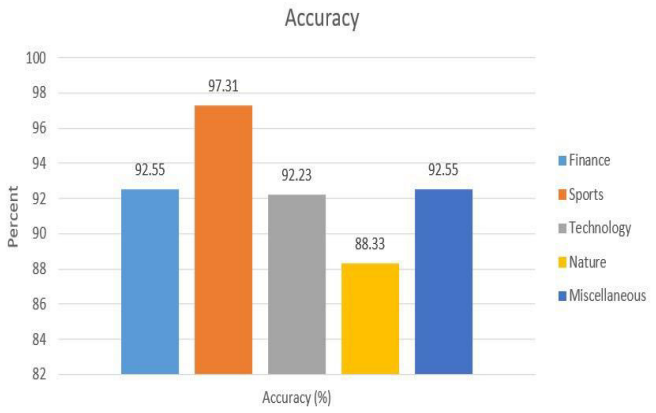

Fig-2: Accuracy

Accuracy represents the proportion of the number of correct samples in the experimental results to the total number of samples. As fig. 2 shows that attribute sport has a high accuracy.

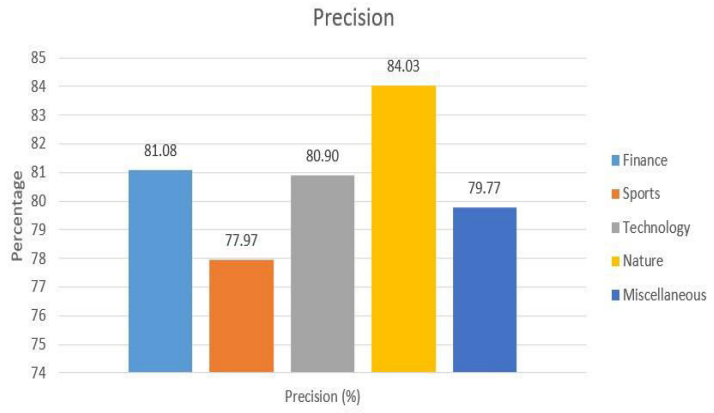

Fig-3: Precision

Precision represents the proportion of the number of correct positive samples in the experimental results to the number of forecast positive sample. As fig. 3 shows that nature has a high precision.

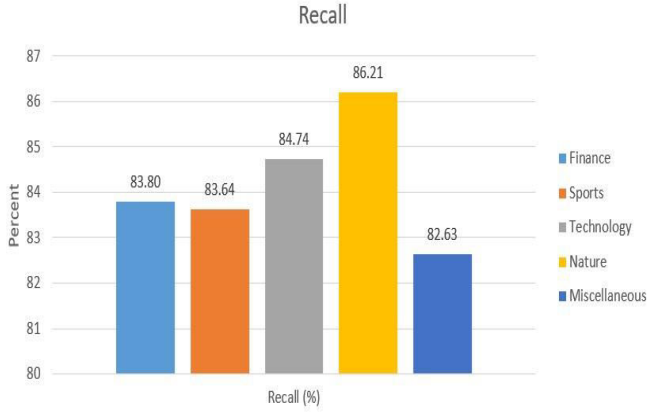

Fig-4: Recall

Recall represents the proportion of the number of correct positive samples in the experimental results to the number of actual positive samples. As show in fig. 4 nature has a high recall value.

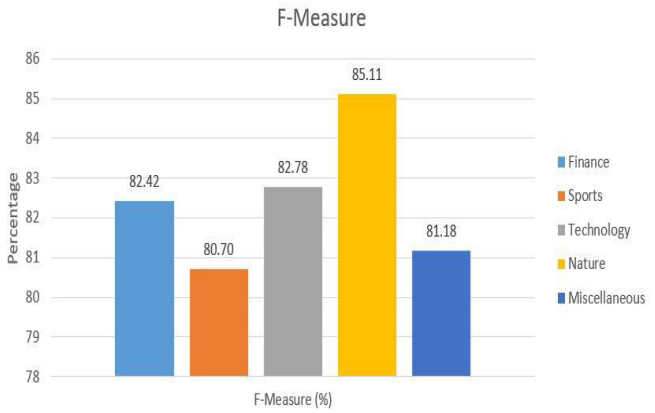

Fig-5:F-measure

F1-measure is the weighted harmonic average of the accuracy rate and recall rate. As show in fig. 5 attribute nature has a high f-measure value.

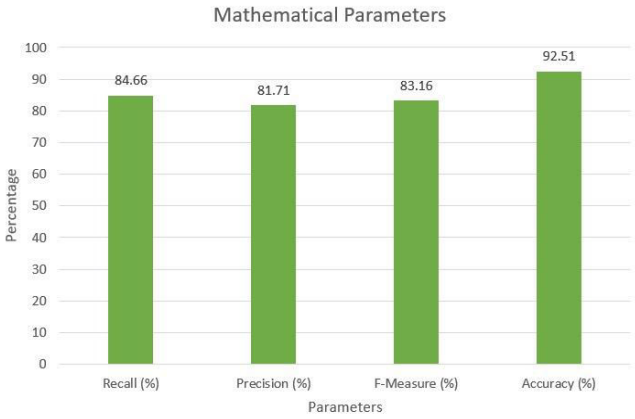

Fig-6: Mathematical Parameter

The overall performance of all the mathematical parameters including Recall, Precision, F-Measure, and Accuracy are consider. As seen in fig. 6 the system has slight less precision but high accuracy.

5. CONCLUSION

The proposed framework give the attribute based access control mechanism. We proposed a method based on support vector machine for attribute mining of unstructured data. Automation mining and extraction of attribute provide the new solution to traditional problem like high cost, poor scalability, maximum manual work, required security expert and time consuming. Attributes are used for access control. SVM is used to recognize and classify the vector attribute, so that the mining of resource attributes can be understand. It provide the less evaluation time. There is no need of security expert and any manual management.

REFERNCES

1. F. Liang, L. H. Yin, and Y. C. Guo, ``A survey of key technologies in attribute-based access control scheme," Chin. J. Comput., vol. 40, no. 7, pp. 1680_1698, 2017.
2. Sandhu RS, Coyne EJ, Feinstein HL, Youman CE. "Role-Based Access Control Models," Computer, 1996, 29(2):38-47.
3. C. Yaokun, Y. Xianglan, and L. Wenli, ``Access control model applicability for big data," Inf. Secur. Technol., vol. 7, no. 7, pp. 3_5, 2016.
4. J. Xin, R. Krishnan, and R. Sandhu, ``A unified attribute-based access control model covering DAC, MAC and RBAC," in Proc. 25th Data Appl. Secur. Privacy, 2012, pp. 41_45.
5. I. Molloy and S. Chari, ``Generative models for access control policies: Applications to role mining over logs with attribution," in Proc. ACM Symp. Access Control Models Technol., 2012, pp. 45_46.
6. L. Yin, F. Liang, B. Niu, B. Fang, and F. Li, ``Hunting abnormal configurations for permission-sensitive role mining," in Proc. Military Commun. Conf., Nov. 2016, pp. 1004_1009.
7. P. P. Griffiths and B. W. Wade. "An Authorization Mechanism for a Relational Database System". ACM Trans. on Database Systems, 1(3):242–255, Sep. 1976.
8. Wenrong Zeng, Yuhao Yang, Bo Luo, "Access Control for Big Data using Data Content," IEEE International Conference on Big Data 2013,pp. 45-47.
9. Xiangwu Ding, Jianming Yang, "An Access Control Model and Its Application in Blockchain," 2019 International Conference on Communications, Information System and Computer Engineering (CISCE),pp. 163-167.
10. Zhongyuan Xu, Scott D. Stoller, "Mining Attribute-Based Access Control Policies from RBAC Policies," 2013 IEEE.
11. Lucas Sterckx, Cornelia Caragea, Thomas Demeester, Chris Develder, "Supervised Keyphrase Extraction as Positive Unlabeled Learning," Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 1924–1929, Austin, Texas, November 1-5, 2016.
12. Mihuandayani, Ema Utami, Emha Taufiq Luthfi, "Text Mining Based on Analysis Using SVM and Feature Selection," 2018 International Conference on Information and Communications Technology (ICOIACT).
13. Yuchun Tang,, Nitesh V. Chawla, "SVMs Modeling for Highly Imbalanced Classification," IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 39, NO. 1, FEBRUARY 2009.
14. Joeran Beel, Akiko Aizawa, Corinna Breitinger and Bela Gipp, "Mr. DLib: Recommendations-as-a-Service (RaaS) for Academia," In Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL), 2017.