Smart Authentication in Music using Triple Genarator

[1]Manjunath R. and [2]Akshatha A.
[2]Associate Professor, Department of Computer Science & Engineering,
City Engineering College, Affiliated to VTU,
Bangalore, Karnataka, India.
[1]Student, M. Tech (CSE), Department of Computer Science & Engineering,
City Engineering College, Affiliated to VTU,
Bangalore, Karnataka, India.

Abstract:- **MUSIC has been an important application area for data mining and machine learning techniques for many years. Music data mining is an interdisciplinary area that studies computational methods for understanding and delivering music data and is a topic of growing importance with large commercial relevance and substantial potential. It attracts researchers not only from computer science, electrical engineering, and musicology but also library science and psychology. The growing popularity and development of data mining technologies bring serious threat to the security of individual's sensitive information. An emerging research topic in data mining, known as privacy preserving data mining (PPDM), has been extensively studied in recent years. The basic idea of PPDM is to modify the data in such a way so as to perform data mining algorithms effectively without compromising the security of sensitive information contained in the data. In particular,we identify four different types of users involved in data mining applications, namely, data provider, data collector, data miner, and decision maker. Management scheme for highly scalable big data mining has not been well studied in spite of the fact that big data mining provides many valuable and important information for us. However, the overlay-based parallel mining architecture is not capable of providing data mining services in case of the physical network disruption that is caused by router/communication line breakdowns because numerous nodes are removed from the overlay network. This paper is about how we manage the music data mining and provide information security using data mining architecture.**

Key Words: Music, Authentication, Data mining Architecture

I. INTRODUCTION

The selected papers underwent a rigorous refereeing and revision process.Efficient and intelligent music information retrieval (MIR) is a very important topic in music data mining. A key step in MIR is to learn feature representations from available music data. Recently unsupervised feature learning techniques (e.g.,sparse coding and deep neural networks) ha e been used to construct audio code words, leading to the bag-of frames (BoF) model and a term-document style representation of music. The paper by Su *et al.* presents a systematic empirical study that compares BoF variants on three MIR tasks (music genre classification, predominant instrument recognition, and audio tagging) by considering different options on codebook learning and encoding, feature pooling, term weighting, power normalization, and dimension reduction. The study improves the understanding of the BoF model and the analogy between traditional text words and audio code words, and also provides several interesting insights in applying the model in MIR. Music recommendation is receiving great attention recently and is becoming prevalent due to the problem of information overload. Its main goal is to provide users with the music pieces that they are likely to enjoy. A good music recommender should have a good balance between relevance (e.g., the recommendations should be similar to the existing relevant items) and novelty (e.g., the recommendations should have some variety and diversity). The paper by Lee and Lee proposes a music recommendation system based on collaborative filtering using dynamically promoted experts to provide recommendations that are novel yet relevant to users. In particular, singular value decomposition (SVD) and K-means clustering are first applied on the user-item matrix to group music pieces into different clusters (or music domains). Then a user is identified as an expert for a music cluster if his/her listening behaviours are concentrated in that cluster. For a target user, different users are selected dynamically as experts for the music clusters on which the user is anovice to provide final recommendations. Experiments demonstrate that the proposed system achieves a good balance between novelty (using music clusters) and relevance (with the help of experts).A challenging problem in music data mining is to model expressive dynamics (e.g., variations in tempo and articulation) and discover expression patterns. Recently many data miningtechniques, in particular, feature learning techniques have beenused to identify useful patterns that are related to the musical expressive performance. Data mining has attracted more and more attention in recentyears, probably because of the popularity of the ``big data" concept. Data mining is the process of discovering interesting patterns and knowledge from large amounts of data.As a high application-driven discipline, data mining has been successfully applied to many domains, such as business intelligence, Web search, scienti_c discovery, digital libraries, etc.With the rapid development of the

information and communication technologies, ``big data'' has been generated from various aspects, such as online transactions, logs, search queries,health records, social networking information, science data, and so forth. It is widely recognized that big data mining is a key component that is required for the actualization of smart society .Since big data comprises various types of data (such as e-mail, social media, video, and sensor data), the big data mining becomes exceedingly complex. Additionally, the big data mining needs to output result data expeditiously in response to the real time demand. Therefore, conventional data mining employsparallel data mining architectures such as Map Reduce and Hadoop to full these requirements.

II. USER ROLE-BASED METHODOLOGY

Current models and algorithms proposed for PPDM mainly focus on how to hide those sensitive information from certain mining operations. However, as depicted in FIGURE, the whole KDD process involve multi-phase operations. Besides the mining phase, privacy issues may also arise in the phase of data collecting or data preprocessing, even in the delivery process of the mining results. In this paper, we investigate the privacy aspects of data mining by considering the whole knowledge-discovery process. We present an overview of the many approaches which can help to make proper use of sensitive data and protect the security of sensitive information discovered by data mining. We use the term ``sensitive information'' to refer to privileged or proprietary information that only certain people are allowed to see and that is therefore not accessible to everyone. If sensitive information is lost or used in any way other than intended, the result can be severe damage to the person or organization to which that information belongs. The term ``sensitive data'' refers to data from which sensitive information can be extracted. Throughout the paper, we consider the two terms ``privacy'' and ``sensitive information'' are interchangeable.

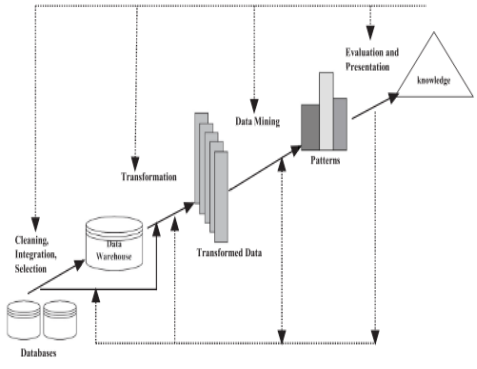

FIGURE 1. An overview of the KDD process.

In this paper, we develop a user-role based methodology to conduct the review of related studies. Based on the stage division in KDD process (see Fig.1), we can identify four different types of users, namely four *user roles*, in a typical data mining scenario (see Fig. 2):

Data Provider: the user who owns some data that are desired by the data mining task.

Data Collector: the user who collects data from data providers and then publish the data to the data miner.**Data Miner**: the user who performs data mining tasks on the data.

Decision Maker: the user who makes decisions based on the data mining results in order to achieve certain goals.In the data mining scenario depicted in Fig. 2, a user represents either a person or an organization. Also, one user can play multiple roles at once. For example, in the Target story we mentioned above, the customer plays the role of dataprovider, and the retailer plays the roles of data collector, data miner and decision maker.By differentiating the four different user roles, we can explore the privacy issues in data mining in a principled way. All users care about the security of sensitive information, but each user role views the security issue from its own perspective. What we need to do is to identify the privacy problems that each user role is concerned about, and to End appropriate solutions the problems. Here we brie_y describe the privacy concerns of each user role. Detailed discussions will be presented in following sections.

1) DATA PROVIDER The major concern of a data provider is whether he can control the sensitivity of the data he provides to others. On one hand, the provider should be able to make his very private data, namely the data containing information that he does not want anyone else to know, inaccessible to the data collector. On the other hand, if the provider has to provide some data to the data collector, he wants to hide his sensitive information

as much as possible and get enough compensations for the possible loss in privacy.

2) DATA COLLECTOR Thedata collected from data providers may contain individuals' sensitive information. Directly releasing the data to the data miner will violate data providers' privacy, hence data modi_cation is required. On the other hand, the data should still be useful after modi_cation, otherwise collecting the data will be meaningless. Therefore, the major concern of data collector is to guarantee that the modi_ed data contain no sensitive information but still preserve high utility.

3) DATA MINER The data miner applies mining algorithms to the data provided by data collector, and he wishes to extract useful information from data in a privacy-preserving manner. As introduced in Section I-B, PPDM covers two types of protections, namely the protection of the sensitive data themselves and the protection of sensitive mining results. With the user role-based methodology proposed in this paper, we consider the data collector should take the major responsibility of protecting sensitive

data, while data miner can focus on how to hide the sensitive mining results from untrusted parties.

4) DECISION MAKE As shown in Fig. 2, a decision maker can get the data miningresults directly from the data miner, or from some *InformationTransmitter*. It is likely that the information transmitterchanges the mining results intentionally or unintentionally,which may cause serious loss to the decision maker. Therefore,what the decision maker concerns is whether the miningresults are credible.In addition to investigate the privacy protection approachesadopted by each user role, in this paper we emphasize a commontype of approach, namely game theoretical approach,that can be applied to many problems involving privacy protectionin data mining. The rationality is that, in the datamining scenario, each user pursues high self-interests in termsof privacy preservation or data utility, and the interests ofdifferent users are correlated. Hence the interactions amongdifferent users can be modelled as a game. By using methodologiesfrom game theory, we can get useful implicationson how each user role should behaviour in an attempt to solvehis privacy problems.

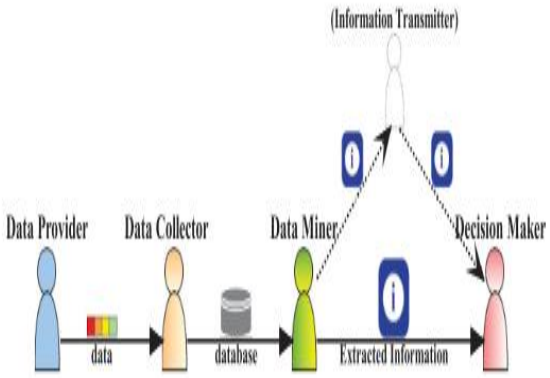

FIGURE 2. A simple illustration of the application scenario with data mining at the core.

III. THE PROCESS OF KDD IN INFORMATION SECURITY

The term ``data mining" is often treated as a synonymor another term ``knowledge discovery from data" (KDD) which highlights the goal of the mining process. To obtainuseful knowledge from data, the following steps are performed in an iterative way (see Fig. 1):

Step 1: Data preprocessing. Basic operations include data selection (to retrieve data relevant to the KDD task from thedatabase), data cleaning (to remove noise and inconsistent data, to handle the missing data FIelds, etc.) and data integration (to combine data from multiple sources).

Step 2: Data transformation. The goal is to transform data into forms appropriate for the mining task, that is, to and useful features to represent the data. Feature selection and feature transformation are basic operations.

Step 3: Data mining. This is an essential process where intelligent methods are employed to extract datapatterns (e.g. association rules, clusters, classication rules, etc.).

Step 4: Pattern evaluation and presentation. Basic operations include identifying the truly interesting patterns which represent knowledge, and presenting the mined knowledge in an easy-to-understand fashion.

IV. OVERLAY-BASED PARALLEL DATA MINING ARCHITECTURE

Overlay-based parallel data mining is one of architecturesthat improve the service availability against server breakdowns. In this architecture, all the servers execute both management and processing functions. The overlay network is constructed by all servers and utilized to End processing nodes, similar to the master nodes in the conventionalarchitecture. This architecture can keep providing the service even if some nodes are removed from the overlay network. Fig. 1 shows an example of mapping and reduction pro- cesses in the overlay-based parallel data mining architecture. When a data processing request is injected, a node that received the request (node A in the Fig. 1) executes reception function by using the overlay network. In other words, the node Ands mappers by using _ooding message

Where mappers are randomly selected (nodes, C, and D in the Fig. 1). Then, a mapper that initially finished the mapping process (node D in the Fig. 1) becomes a reducer, and it requests to other mappers to transmit the processed data to itself, where the request message can be forwarded by using _ooding scheme. After receiving the processed data from mappers, the reducer executes the reduction process and outputs the analysed result.In this architecture, since the connectivity of overlay network dramatically affects the service availability ofdata mining, there are numerous works, which tackled the connectivity issue from the various viewpoints, i.e., context aware, graph theory based, and complex network theory based overlay network construction schemes [21]_[24]. These works make overlay networks that are tolerant to small-scale server breakdowns but do not consider the large-scale server breakdowns, i.e., physical network disruption. Therefore, this paper develops an overlay-based parallel data mining architecture that is tolerant to physical network disruption so that data mining is available at anytime, anywhere.

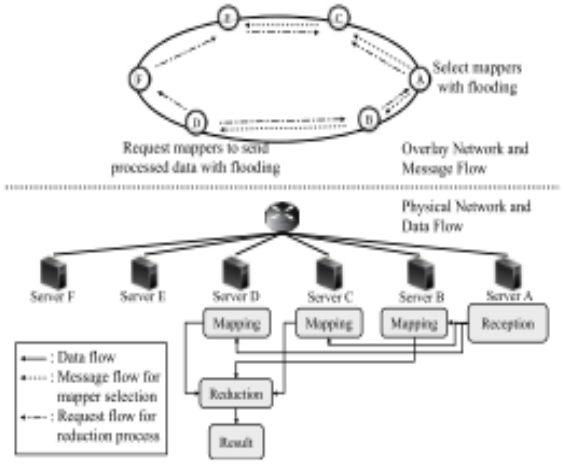

FIG FOR PARALLEL DATA MINING ARCHITECTURE

V. CONCLUSION

Music is of rich structure with its hierarchical and multi-faced organization of basic elements. Hence inferring and annotating the structural properties of music pieces is an important problem in music data mining as they play important roles in music indexing, similarity search and information retrieval. In their paper, Smith, Chew, and Chaun investigate a large corpus ofannotated recordings and study the relationship between the acoustic properties of recordings and the boundary indications of listeners. Their results demonstrate a strong correlation between the acoustic boundaries, estimated by musical features reflecting timbre, harmony, key, rhythm and tempo, and the annotated boundary positions and the strength of the relationship is moderately affected by the musical feature representation. In another work, Serra *et al.* present an unsupervised approach based on structure features and time series similarity to automatically detect the temporal locations of segment boundaries (i.e., novelty and boundary detection) and to assess segment similarities and repetitions (i.e., classification/grouping). The structure features jointly consider both local and global aspects by measuring the relations between each time frame/window with all other times frames/window in a music piece and they are used to obtain reliable segment boundaries. The obtained segment boundaries are then used for music structure labelling based on time series similarity measures.To achieve the privacy-preserving goals of different users roles, various methods from different research FIelds are required. We have reviewed recent progress in related studies, and discussed problems awaiting to be furtherinvestigated. We hope that the review presented in this paper can offer researchers different insights into the issue of privacy-preserving data mining, and promote the exploration of new solutions to the security of sensitive information.An overlay-based data mining architecture, which fully distributes management and processing functions by using overlay network technologies, can potentially provide scalable data mining in large-scale network. However, due to physical network disruption, this architecture dramatically decreases service availability of data mining. To solve this problem, we proposed neighbour selection and task allocation schemes based on integration of the overlay and physical networks. In order to improve the success probability of data mining against physical network disruption, our neighbour selection scheme constructs overlay network based on node location in physical network and our task allocation scheme selects nodes from different diagonally-cornered groups in the overlay network as mappers. Moreover, the results obtained from the numerical analysis demonstrated the effectiveness of our proposed schemes in terms of sign cant improvement in the service availability. Thus, our proposed schemes can be

VI. REFERENCES

[1]J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. San Mateo, CA, USA: Morgan Kaufmann, 2006.
[2]L. Brankovic and V. Estivill-Castro, ``Privacy issues in knowledge discovery and data mining," in *Proc. Austral. Inst. Comput. Ethics Conf.*, 1999, pp. 89_99.
[3]R. Agrawal and R. Srikant, ``Privacy-preserving data mining," *ACM SIGMOD Rec.*, vol. 29, no. 2, pp. 439_450, 2000.
[4]Y. Lindell and B. Pinkas, ``Privacy preserving data mining," in *Advancesin Cryptology*. Berlin, Germany: Springer-Verlag, 2000, pp. 36_54.
[5]C. C. Aggarwal and S. Y. Philip, *A General Survey of PrivacyPreserving Data Mining Models and Algorithms*. New York, NY, USA: Springer-Verlag, 2008.
[6]M. B. Malik, M. A. Ghazi, and R. Ali, ``Privacy preserving data mining techniques: Current scenario and future prospects," in *Proc. 3rd Int. Conf.Comput. Commun. Technol. (ICCCT)*, Nov. 2012, pp. 26_32.
[7]S. Matwin, ``Privacy-preserving data mining techniques: Survey and challenges," in *Discrimination and Privacy in the Information Society*. Berlin, Germany: Springer-Verlag, 2013, pp. 209_221
[8]E. Rasmusen, *Games and Information: An Introduction to Game Theory*, vol. 2. Cambridge, MA, USA: Blackwell, 1994.
[9]V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati, ``Microdata protection," in *Secure Data Management in DecentralizedSystems*. New York, NY, USA: Springer-Verlag, 2007, pp. 291_321.
[10]O. Tene and J. Polenetsky, ``To track or `do not track': Advancing transparency and individual control in online behavioral advertising," *Minnesota J. Law, Sci. Technol.*, no. 1, pp. 281_357, 2012.
[11]R. T. Fielding and D. Singer. (2014). *Tracking Preference Expression (DNT). W3C Working Draft*. [Online]. Available: http://www.w3.org/TR/2014/WD-tracking-dnt-20140128/
[12]R. Gibbons, *A Primer in Game Theory*. Hertfordshire, U.K.: Harvester Wheatsheaf, 1992.
[13]D. C. Parkes, ``Iterative combinatorial auctions: Achieving economic and computational ef_ciency," Ph.D. dissertation, Univ. Pennsylvania, Philadelphia, PA, USA, 2001.
[14]S. Carter, ``Techniques to pollute electronic pro_ling," U.S. Patent.
[15] T. Li, M. Ogihara, and G. Tzanetakis, *Music Data Mining*,ser. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. Boca Raton, FL, USA: CRC Press, 2012.
[16] T. Li and M. Ogihara, "Toward intelligent music information retrieval," *IEEE Trans. Multimedia*, vol. 8, no. 3, pp. 564–574, Jun. 2006.