RESEARCH ARTICLE OPEN ACCESS

Optimization Technique of Association with ACO for High Resolution Image Classification: Survey

Deepika Goyal[1], Shruti Grover[2]
[1]Research Fellow, [2]Asst. Professor
Department of Computer Science and Engineering
Shri Guru Granth Sahib World University.
[1]goyaldeepika4@gmail.com, [2]shruti_cse@sggswu.org

ABSTRACT
Data mining is a process of discovering patterns and relationships in data with the help of various data analysis tools, to make valid predictions. Association rule learning which finds the relationships between the variables. Association rules are important features for image classification, mining and rational selection to obtain accurate classification. In this paper, it is an approach to present association rules analysis & ACO together and applied to object-oriented of high resolution image classification. Sometimes association rule is referred to as market basket analysis. Classification is the task of classify the new data by using classification algorithm. In this paper I work on the implementation of optimization algorithm (ACO) based on artificial intelligent scheme for finding more optimal rules for classification.

Keywords- Association rule, Apriori algorithm, ACO (Ant Colony Optimization Algorithm), Classification of high resolution image

I. INTRODUCTION

Data mining is the important tool that is used for extracting the data from the large amount of data base. Data is stored in the data warehouse, relational database and the transactional database. . The first and simplest analytical step in data mining is to describe the data — summarize its statistical attributes (such as means and standard deviations. As emphasized in the section on the data mining process, collecting, exploring and selecting the important and correct data are very important. Classification and association rules are two important techniques of data mining in knowledge mining process. Integration of these two techniques has produced new approaches called Associative Classification Technique or Class Association Rule Mining. Classification is one of the most important tasks in data mining. With the development of technology, the high resolution of image has increased enormously for the past decades. Image classification is one of the main methods for information extraction from high resolution of image. There are many classification approaches for extracting knowledge from data such as statistical [1], divide-and-conquer [2] and covering [3] approaches. Numerous algorithms have been derived from these approaches, such as Naiave Bayes [1], See5 [4], C4.5 [5], PART [6], Prism [6] and IREP. However, traditional classification techniques often produce a small subset of rules, and therefore usually miss detailed rules that might play an important role in some cases [7]. In this paper Apriori algorithm used for association rule. Apriori is a classification algorithm that is used for mining frequent item set. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. Apriori algorithm can be used to determine association rules which highlight general trends in the database for finding the frequent item set: this application is used in domains such as market basket analysis.

The remainder of this paper is organized as follows. In section 2 we briefly mention association rules mining algorithm, Section 3 Apriori algorithm, Section 4 Classification of high resolution image, Section 5 ACO, Section 6. Apriori algorithm with ACO, Section 5 Conclusion and Future work.

II. ASSOCIATION RULE MINING

A. Association Rules

In this paper Association rule learning which finds the relationships between variables. Association rule mining has become an important data mining technique due to the descriptive and easily understandable nature of the rules. For example a market might gather data on customer bought things. Using association rule learning, the market can determine which products are frequently bought together. Although association rules mining was introduced to extract associations from market basket data. it has proved useful in many other domains (e.g. micro array data analysis, recommender systems, and network intrusion detection). A

common way of measuring the usefulness of association rules is to use the support-confidence. The confidence is the percentage of take all the items in the rule among those transactions that carry the items and Support of a rule is the percentage of transactions that carry all the items in the rule. Most of association rules mining algorithm are based on the Apriori property [8], which means that all non-empty subsets of a frequent item set must also be frequent.

B. SUPPORT AND CONFIDENCE

Generate all association rules from the set of transactions that have support greater than or equal to min support and confidence greater than or equal to min confidence; The problem of association rule mining can be stated as: Given a dataset of transactions, a threshold support (min support), and a threshold confidence (min confidence). Let I be a set of items, D be a data set containing transactions (i.e., sets of items in I) and t be a transaction. An association rule mined from D will be of a form X -> Y, where X; Y C I and X intersection Y =Φ The confidence is, out of all the transactions that contain X, the percentage that contain Y as well. The support of the rule is the percentage of transactions in D that contain both X and Y. The confidence of a rule measures the strength of
the rule (correlation between the antecedent and the consequent) while the support measures the frequency of the antecedent and the consequent together.

III. APRIORI ALGORITHM

A. APRIORI ALGORITHM

The Apriori algorithm was introduced for generating association rules. The Apriori algorithm is a two stage process: a rule generation stage (rules that satisfy minimum confidence threshold) and frequent item set (item sets that satisfy minimum support threshold) mining stage. Apriori algorithm is used for frequent item set .Apriori algorithm can be used to determine association rules which highlight general trends in the database for frequent item set: this application is used in domains such as market basket analysis. . It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. Each transaction is seen as a set of items (an item set). Given a threshold C, the Apriori algorithm identifies the item sets which are subsets of at least C transactions in the database. Other algorithms are designed for finding association rules in data having no transactions (Winepi and Minepi), or having no timestamps (DNA sequencing). Apriori is designed for working on databases that containing various transactions (for example, collections of items purchased by customers). Apriori uses a "bottom up" approach where frequent subsets are extended one item at a time that is called candidate generation. Apriori uses breadth-first search and a Hash tree structure for efficiently count candidate item sets,. It generates candidate item sets of length k from item sets of length $k-1$.

In Table 3.1, we show a subset of the contact-lenses data set from the University of California Irvine (UCI) Machine Learning Repository .We will generate association rules from this data set (Figure 3.1). Each attribute-value pair is referred to as an item. For brevity, attribute-value pairs are denoted only by their values. For example, age = young will be written as young.

Table 3.1

Age	astigmatism	tear-prod-rate	Contact-lenses
Young	No	Normal	Soft
Young	Yes	Reduced	None
Young	Yes	Normal	Hard
Pre-presbyopic	No	Reduced	None
Pre-presbyopic	No	Normal	Soft
Pre-presbyopic	Yes	Normal	Hard
Pre-presbyopic	Yes	Normal	None
Presbyopic	No	Reduced	None
Presbyopic	No	Normal	None

B.THE FREQUENT ITEMSET MINING STAGE

In the first iteration of Apriori's frequent item set mining stage, each item becomes part of the 1 item candidate set C1. The algorithm is used for the data set to count support for C1 in fig 3.1 Those item sets satisfying the minimum support will form L1, the set of frequent item sets of size 1.. The ones that have support less than the minimum support threshold are shown in gray in Figure 3.1. To generate candidates of size 2 (C2) item sets, the level 1 collection of frequent item sets is joined with itself. This join is denoted by L1 on L1 and is equal to the collection of all set unions of different item sets in L1. The algorithm scans the database for support of the items in C2. Those item sets satisfying the minimum support condition will form L2. When generating

candidates of size 3 (C3), L2 on L2 is performed but with a condition. Apriori assumes that the items in an item set are sorted according to a predefined order (e.g. lexicographic order). The join, L_k on L_k for k > 1, has the condition that for two item sets from L_K to be joined, the first k 1 item(s) must be the same in both item sets. This ensures that the generated candidate is of size k and that most of the subsets of the set are frequent. The Apriori property is applied, before counting support for all the items in C3,. The Apriori property states that all nonempty subsets of an item set must be frequent for any given item set .The Apriori property prunes the search space. The Apriori algorithm continues to generate frequent item sets until it cannot generate any more candidate item sets.

IV. CLASSIFICATION OF HIGH RESOLUTION IMAGE

The classification algorithm learns from the training set and builds a model and that model is used to classify new objects. Classification and prediction are two forms of data analysis that can be used to extract important data classes. This type of analysis helps us for understanding of the data at large scale. The data analysis task is classification; Decision tree induction is the learning of decision trees from class-labeled training tuples. The pixel-based image classification can't satisfy high-resolution satellite image's classification precision. Object-oriented information extraction also uses geometry and structure information not only depends on spectrum character. Increasing rate of multimedia data, remote sensing and web photo gallery need a category of different image for the proper retrieval of user. Image classification is vital field of research in computer vision. The proper selection of feature are challenging task of classification, so various authors apply some machine learning approach for image classification such as decision tree, RBF network, Markova model and support vector machine. In this paper we review of machine learning approach for image classification. Various researchers apply different approach for image classification such as segmentation, clustering and some machine learning approach for the classification of image. High-resolution satellite images offer abundance information of the earth surface for remote sensing images. The information includes attribute characteristic, geometry texture and color, texture and shape and size plays an important role in image classification. We examine the use of high frequency features in high resolution images and demonstrate how they can increase texture classification accuracy when used in combination with lower frequency features. We used eight features, four low frequency and four high frequency. This research is to extract a kind of "structure" from a sample of objects. This paper discusses one of the most widely used supervised classification techniques with ACO and evaluate strength of own classification with Performance analysis and Results analysis. Data classification means categorization of data into different category according to rules.. To rephrase it better to learn a concise representation of these data.

V. ANT COLONY OPTIMINZATION

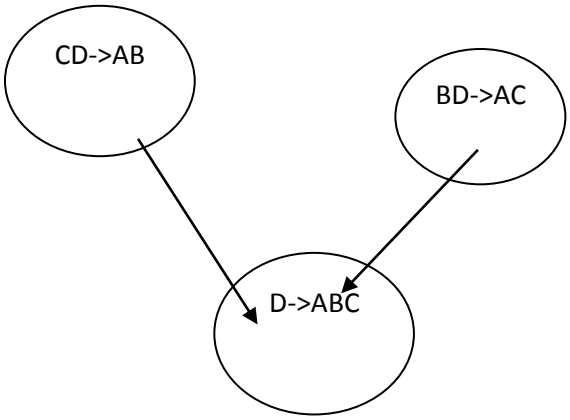
The intelligent behaviour of the colony emerges from the indirect communication between the ants mediated by small modifications of the environment, which is called stigmergy.ACO is a self-organised systems which can accomplish complex tasks by having their individual ants interacting with one another and with their environment. Ants drop pheromone on the ground as they walk from a food source to the nest by using the path. Many ant species are able to find the shortest path between a food sources. Given that shorter paths are traversed faster than longer ones, they have a stronger pheromone concentration after a period time, contributing to being selected and forcefully applied. Colony of artificial ants are used by ACO algorithms where ants build candidate solutions to optimization problems by iteratively selecting solution components based on their associated pheromone and heuristic information—where the latter corresponds to a measure of how good a solution component is for the problem at hand. In ACO algorithms, artificial ants are used for construct the candidate solutions by using traversing a graph, called the construction graph. This is a graph where each vertex represents a potential component of a candidate solution, and the act of traversing an edge means that an ant is adding, to the current candidate solution, the component vertex at the end of the edge. Dorigo et al. [14][15] [16]have defined an artificial ant colony meta heuristic that can be applied to solve optimization problems, called Ant Colony Optimization (ACO).

VI. APRIORI ALGORITHM WITH ACO (Ant Colony Optimization Algorithm)

Association rules are used the apriori algorithm for mining frequent item set. Rule Genration for apriori algorithm in which having candidate rule is generated by merging two rules that share the same prefix in the rule consequent.

Join (CD->AB,BD->AC) would produce the candidate rule D->ABC if does not have high confidence. Support count has been obtained during the frequent item set generation step. Apriori algorithm is reduced the number of comparison by scan the database of transactions to determine the support of each generated candidate item set. To reduce the number of comparison, store the candidates in a hash structure. Instead of matching each transaction against every candidate, match. It against candidate contained in the hashed buckets.

S



In this paper we propose a novel ACO algorithm to induce with apriori, combining commonly used strategies from both traditional with apriori and ACO. Ant Colony Optimization (ACO) algorithms have been successfully applied to extract classification rules; Apriori with ACO algorithms remains an almost unexplored research area. The colony produces a system capable of performing a robust search to find high-quality solutions for optimization problems with a large search space. In the context of the classification task in data mining, ACO algorithms have the advantage of performing a flexible robust search for a good combination of item set, less likely to be affected by the problem of attribute interaction [9, 10]. Ant colony optimization (ACO) [11, 12, 3] algorithms involve a colony of ants (agents), which despite the relative simplicity of their individuals' behaviours, cooperate with one another to achieve an unified intelligent behaviour.

VII. CONCLUSION AND FUTURE WORK

In recent years, data-mining (DM) has become one of the most valuable tools for extracting and manipulating data and for establishing patterns in order to produce useful information for decision-making. The increasing size of multimedia databases and the ease of accessing them by a large number of users through the Internet carry a problem of efficient and semantically adequate querying of such content. Querying of images by content is an important part of any multimedia mining system.

Image are generated at increasing rate by sources such as military reconnaissance flights; defence and civilian satellites; fingerprinting devices and criminal investigation; scientific and biomedical imaging; geographic and weather information systems; stock photo databases for electronic publishing and news agency; fabric and fashion design; art galleries and museum management; architectural and engineering design; and WWW search engines. Most of the existing image management systems are based on the verbal descriptions to enable their mining. Image mining is then based on standard mining. Image mining presents special characteristics due to the richness of the data that an image can show.

In previous work, an approach that integrates association rules analysis and ACO is presented and applied to object-oriented high resolution image classification. The association rules analysis is adopted for mining strong rules from an image, and the ACO is for finding the optimal rules for classification in artificial intelligence. In our proposed work, we will implement optimization algorithm (ACO) based on artificial intelligent scheme for finding more optimal rules for classification with association rule.

REFERENCES

[1] John, G. H. and Langley, P. (1995). *Estimating Continuous Distributions in Bayesian Classifiers*. Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. pp. 338- 345. Morgan Kaufmann, San Mateo.

[2] Furnkranz, J. (1996). *Separate-and-conquer rule learning*. *Technical Report* TR-96-25, Austrian Research Institute for Artificial Intelligence, Vienna.

[3] Cendrowska, J. (1987). *PRISM: An algorithm for inducing modular rules*. International Journal of Man-Machine Studies. Vol.27, No.4, pp.349-370.

[4] Quinlan, J. R. See5.0 (http://www.rulequest.com) viewed on May 2010.

[5] Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, San Francisco.

[6] Frank, E. and Witten, I. (1998). Generating accurate rule sets without global optimization. In Shavlik, J., ed., *Machine Learning: Proceedings of the Fifteenth International Conference*, pp.
144-151, Madison, Wisconsin, Morgan Kaufmann, San Francisco.

[7] Pazzani, M., Mani, S., and Shankle,W. R. (1997). *Beyond Concise and colorful: learning intelligible rules*. In *KDD-97*.

[8] R.Agrawal, T. Imielinski and A. Swami. "Mining association rules between sets of items in large databases," Proc. of the ACM SIGMOD Conf. on Management of data, pp. 207-216, 1993.

[9] V. Dhar, D. Chou, and F. Provost. Discovering Interesting Patterns for Investment Decision Making with GLOWER – A Genetic Learner Overlaid With Entropy Reduction. *Data Mining and Knowledge Discovery*, 4(4):251–280, 2000.

[10] A.A. Freitas. Understanding the crucial role of attribute interaction in data mining. *Artificial Intelligence Review*, 16(3):177–199, 2001.

[11] M. Dorigo and G. Di Caro. The Ant Colony Optimization meta-heuristic. InD. Corne, M. Dorigo, and F. Glover, editors, *New Ideas in Optimization*, pages 11–32, 1999.

[12] M. Dorigo, G. Di Caro, and L.M. Gambardella. Ant algorithms for discrete optimization. *Artificial Life*, 5(2):137–172, 1999.
[13] M. Dorigo and T. St utzle. *Ant Colony Optimization*. MIT Press, 2004.
[14] M. Dorigo and G. Di Caro. The Ant Colony Optimization meta-heuristic. In D. Corne, M. Dorigo, and F. over, editors, *New Ideas in Optimization*, pages 11–32, 1999.
[15] M. Dorigo, G. Di Caro, and L.M. Gambardella. Ant algorithms for discrete optimization.
Artificial Life, 5(2):137–172, 1999.
[16] M. Dorigo and T. St ¨utzle. *Ant Colony Optimization*. MIT Press, 2004