

## **ACO based Mining of a Cloud Data Center**

Vipul Snehadeep Chawathe<sup>1</sup>

<sup>1</sup>VJTI, Student, H R Mahajani Marg, Matunga, Mumbai- 400019

---

**Abstract**— *The ACO algorithms belong to Swarm family of Intelligent Systems in Artificial Intelligence domain. They are natural choice for the task of data mining when the source can be logically partitioned. A cloud host data center is seen partitioned as such amongst the clouds. As clouds are deployed as per SLAs, the data spread across different clouds serves the service provider in furthering the scope for providing business services.*

**Keywords**— *Intelligent Systems, Swarm, Cloud, Mining, KDD*

---

### **I. INTRODUCTION**

Ant Colony Optimization (ACO) was inspired by the behaviour of ants and has many successful applications in discrete optimization problems. An artificial ant colony system (AACS) is a random stochastic population based heuristic algorithm of agents that simulate the natural behaviour of ants, developing mechanisms of cooperation and learning, which enables the exploration of the positive feedback between agents as a search mechanism[1].

Some ACO mining algorithms have been proposed to perform the classification task within the data mining operation for Knowledge Discovery in Databases. Such algorithms typically give the computational theory language rules for the data being classified. Data classification is useful when preparing search indexes whereas some application might require data access rather than search which will use clustering. Herein only classification is the target.

A data center may host data warehouse, or many data marts. The sheer vastness of data present might overwhelm normal persons to operate with it. Thus, it is a suitable candidate for automated mining. Unlike consumer software, operating such computing facility seldom fails to be tackled using Object Oriented paradigm. The ants approach is an agent-based approach.

Cloud hosting is where data centers really shine in electronic media. Every major IT firm has an offering of how consumers can be served by data centers wherein their cloud will be scaled as per Service Level Agreement (SLA) bounded only by the capacity of the data center. Hence, almost any cloud related product offering for cloud services providers will offer to visualize logical clubbing amongst the data centers internal resources. By mining across these cloud boundaries the service provider should be better prepared for upcoming business.

The remaining paper is organized as following. Section II covers the ant mining algorithms that are evolved into designs from research. The III Section looks at how data can be made available from the cloud. Wide variety of people has always been served by data centers, it's just that mainstream commercialization has required its standardization now so we've terms such as clouds & SLAs. Owing to the standardization being recent this part is more seminal. Section IV concludes with some data for classification.

### **II. RELATED WORK**

Ant mining algorithms have been around for almost two decades. The first AntMiner was purely aimed at simply performing the mining task by using swarm, which was defined in late eighties, as compared to its later versions. The later versions improvised the performance by utilizing knowledge of various parameters. Current trend is towards hybrid algorithms.

The first algorithm for consideration will be Ant-Miner (ant-colony-based data miner)[2]. It introduced the role of heuristic function and pheromone updating strategy to the mining domain. In a simple ACO algorithm, the main task of each artificial ant is to find a shortest path between a pair of nodes on a graph on which the problem representation is suitably mapped. Each ant applies a step-by-step constructive decision policy to build the problem's solution. This may be collectively referred as Probabilistic Transition Rule or as considered herein- the heuristic function.

The second algorithm that's noteworthy is the AntMiner+[3]. While the Ant-Miner had improvised versions such as AntMiner3 that improved in the heuristic function, AntMiner+ improvised classification by

- Environment:

- The environment is defined as a directed acyclic graph (DAG), so that the ants can choose their paths more effectively.
- To allow for interval rules, the construction graph additionally exploits the difference between nominal and ordinal variables.
- Inclusion of the weight parameters for the pheromone and heuristic value in the construction graph.
- Implementation of the better performing MAX-MIN Ant System.
- The usage of more accurate class-specific heuristic values, with the exclusion of the majority class to be used as the final default class.
- Application of an early stopping criterion.

The third noteworthy algorithm is the Grammar Based Ant Programming (GBAP[4]). It's guided by a CFG but uses ACO as a search technique. The GBAP algorithm is conceived for obtaining a specific classifier arising from a learning process over a given training set. GBAP initializes all edges to maximum pheromone level allowed and limits the pheromone level to some particular range. So it's impure.

### III. GENERAL DESCRIPTION OF DATA RETRIEVAL FROM WITHIN CLOUD

The cloud consumer typically gets bunch of virtual machines hosted by a data center so they may be scaled as per services demands. In such context the term "service" means something which is available as per demand, which is typically subjected to abrupt variations. So we first need to understand the structure of service. This should be followed by capabilities of the virtual machines. Finally, we are ready for data retrieval. This is covered by the following subsections.

#### A. Structure of Cloud Service

Services are typically categorized as:

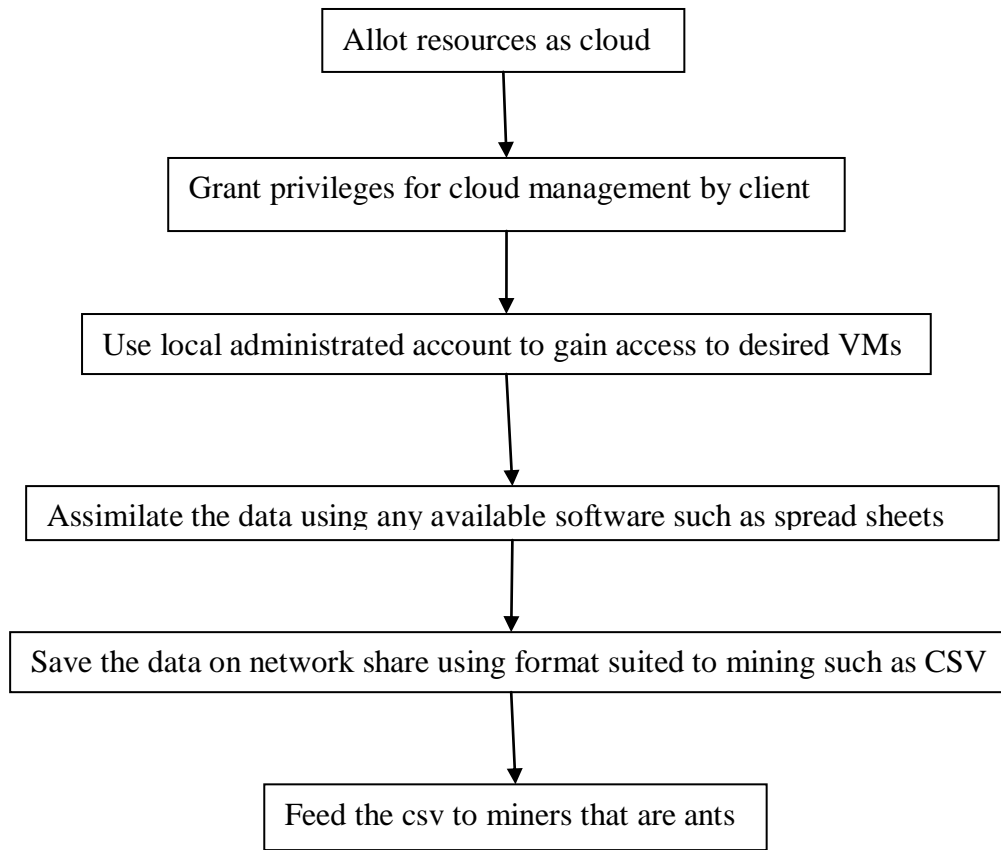
Service	Description
SaaS	Software as a Service provides remote application hosting. It allows doing things such as storing photo albums online, so they can be accessed by anyone authorized within the uploader's group.
IaaS	Infrastructure as a service means the software might require memory as well as processor resources, such as multi-tier web hosting that will have separate virtual machines (VMs) for backend database and front-end web hosting.
PaaS	Platform as a Service an OS image is maintained, so client can configure usage instances. For eg. Gmail for businesses, where the user-ids are set by the organization.
Haas	Hardware as a Service gives remote control over raw hardware resources. The client user might just require allotted CPU count to be increased.

Table 1: Cloud Service Categories Description

Observe that in all services remote control operation is present, be it for end user or administration by client. The resources thus handled are easily broken into units of virtual machines. Now for data related operation we may set aside the last form, viz. HaaS. Also observe that as all interaction is with these VM units the fabric backbone is the logically virtual network part. As user operation is authenticated, some domain user system is typically setup, similar to Active Directory on Windows systems.

#### B. Capabilities of Virtual Machines

Now it can be understood that for virtual machine existence some virtual machine manager (VMM) layer is present over the host data center OS. The next thing needed is to access the VM OS through VMM. This might be as simple as accessing VM using local user account of guest OS in Citrix Xen or VMWare or Hyper-V VMM. From such a connection all that remains to be done is to send files outside the system, which may require sending email outside as an extreme case or might be as straightforward as copying file to some shared network location. As these are commercial tools subject to licensing, the working is exhibited by block diagram rather than screenshots.



*Fig. 1* A Workflow of actions performed by cloud data center authority for mining

### C. Knowledge Discovery of Classes of Data

By allowing ants to wander through the CSVs, classification will give insights such as:

- 1) **Usage Details:** Users might have different types of documents motivating the Cloud Service Provider to supply finer word processing software.
- 2) **Timing Profile:** A Depending on when the VM becomes less responsive due to maybe automated maintenance or usage traffic, resource scaling to the cloud may be optimized by the CSP.

Default choice for modelling ants is the functional programming construct of agents. Agents are reactive by nature. An ant, either by incrementing pheromone or letting it evaporate, reacts to every node of data it visits. In the simplest case this might be an attribute within the CSV file. Also proper use of functional programming guarantees asynchronous parallelism, concurrency.

The ants update pheromone values every time they pass through nodes. These are available for future ants as reference. Also the previously obtained pheromone readings can be continued in use even as new data becomes available. Two ants never interact with each other in anyway besides pheromone reading reuse. This is asynchronous operation.

While one is reading data, another might be updating pheromone. This guarantees concurrency.

Multiple ants might be applying probabilistic transition rule at different node. This is parallel operation.

Such asynchronous, parallel, concurrent, i. e. agent based, mining architecture is significant as the purpose of data center is not mining but serving the cloud. If such operation is unavailable then the ants' population will cause a denial of service (DoS) attack to the hosted clouds! Also recap from the beginning why intelligence was required and how much of data was expected in data centers. This enables to setup the concluding proposal

#### IV. CONCLUSIONS

While multiple approaches to Knowledge Discovery in Databases are present, ACO based mining of Cloud Data Center has a strong logical appeal for the case in point. The architecture of cloud hosts renders nicely with swarm operations as finally explained in the application of ants over foddered data. Swarm functionality implemented through functional agents will be light-weight for the data centers in question. Although there is deviation from mainstream trend of applying object oriented paradigm, functional paradigm should be further explored on data centers for mundane implicit tasks such as clouds mining.

#### V. ACKNOWLEDGMENT

The author wishes to acknowledge his guide Ravita Mishra for the training on data mining imparted by her. The author also expresses his gratitude to Bandu Meshram for actively nourishing the authors interest in clouds.

#### REFERENCES

- [1]. M. Dorigo, *The ant system: Optimization by a colony of cooperating agents*, IEEE Transactions on Systems, Man, and Cybernetics-Part B 26(2): 29-41.
- [2]. Rafael S. Parpinelli, Heitor S. Lopes, Alex A. Freitas, *Data Mining with an Ant Colony Optimization Algorithm*, IEEE Transactions on Evolutionary Computing, Vol. 6 No. 4, August 2002.
- [3]. David Martens, Manu De Backer, Raf Haesen, Jan Vanthienen, Monique Snoeck, Bart Baesens, "Classification with Ant Colony Optimization" *IEEE Transactions on Evolutionary Computation*", vol. 11, No. 5, October 2007.
- [4]. Juan Luis Olmo, Jose Raul Romero, Sebastian Ventura, "Using Ant Programming Guided by Grammar for Building Rule-Based Classifiers" in IEEE Transactions on Systems, Man, and Cybernetics- Part B: Cybernetics, Vol. 41, No. 6, December 2011.