Chapter 10
An Intense Study on Intelligent Service Provisioning for Multi-Cloud Based on Machine Learning Techniques

C. D. Anisha and K. G. Saranya

10.1 Introduction

The introduction section provides overview on services of cloud computing and need, terminologies, and challenges of multi-cloud computing.

10.1.1 Cloud Computing and Its Services

Cloud computing is a term that represents a distributed model which allows sharing of resources in form of services without any restriction of time (anytime), location (anywhere), and user (anyone). This environment provides the services to users in pay as you go manner, i.e., on a subscription basis. The basic requirement to use cloud services is Internet connectivity. The services related to education, business, and governance are provided through online wherein users can access though web browser. The most prominent cloud service providers are Amazon, Google, and Microsoft [1]. The datacenter which consists of cloud server that has data and software programs stored in it provides resources to any user in the world. The approach of cloud computing enables effective use of resources and to procure and update data of a user without any purchasing of license.

C. D. Anisha (✉)
Research Scholar (Ph.D), Department of CSE, PSG College of Technology, Coimbatore, India

K. G. Saranya
Assistant Professor (Sr.Gr), Department of CSE, PSG College of Technology, Coimbatore, India
e-mail: kgs.cse@psgtech.ac.in

R. Nagarajan et al. (eds.), *Operationalizing Multi-Cloud Environments*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-030-74402-1_10

10.1.2 Multi-Cloud Computing: Need, Terminologies, and Challenges

Multi-cloud is referred to as cloud system in which applications are hosted in blocks among a diverse network of distinct cloud.

The need for multi-cloud is as follows:

1. *Choice of service*: Single cloud service provider (CSP) vendor will not provide all services essential for the organization. Integrating utilization of services from multiple CSP is made possible through multi-cloud.
2. *Latency*: The service can be chosen based on the distance. The service which is nearer to the user can be provided in multi-cloud environment.
3. *Increased disaster recovery*: Multi-cloud environment allows the replicas of applications in two or more clouds. This capability allows the user to access another cloud if there is downtime in one cloud.

Terminologies and its associated tools used in multi-cloud are as follows:

1. *Library-based approaches:*
   (a) *jclouds*: It is an open-source java library used to provide mobility of java applications.
   (b) *libcloud*: It is a python library that provides detachment of various differences among the programming interfaces.
2. *Service-based approaches:*
   (a) *Hosted services*: These are commercial services, and the most commonly hosted services are RightScale, enStratus, and Kaavo.
   (b) *Deployable services*: These are open-source projects, and the most commonly used deployable services are Aeolus, mOSAIC, and optimis.

Challenges in multi-cloud are as follows:

1. *Cost management*: The biggest challenge in multi-cloud is the management of cost, as different cloud service providers (CSP) fix price with different meters. The construction of multi-cloud with available cost in organization is a huge challenge in multi-cloud.
2. *Assets management*: Organization will get access from developed CSPs and developing CSPs. Complexity of managing asset increases when integrating the developed and developing CSPs in multi-cloud.
3. *Performance management*: Organization performance will be affected if the services are chosen in a substandard manner.
4. *User management*: Managing access rights among users becomes a challenge in multi-cloud environment.

10.2 Classification of Services Provisioning in Multi-Cloud

Service provisioning is the process of allocating the on-demand resources to the user. The service provisioning provides four major services, namely, SaaS (Software as a Service), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Security as a Service (SecaaS). Service provisioning involves monitoring and orchestration.

The service provisioning in multi-cloud can be classified based on the following:

1. The characteristics, namely, workload, elasticity, and location (placement and consolidation)
2. The approaches for service provisioning:
   (a) Service provisioning models
   (b) Brokerage aided provisioning
   (c) Policy ensure service level agreements (SLAs)
   (d) Heuristic and holistic perspective
   (e) Multi-criteria decision-making (MCDM)
   (f) Algorithmic techniques
3. Services orchestration at each service level:
   (a) IaaS service orchestration
   (b) PaaS service orchestration
   (c) SaaS service orchestration

10.2.1 Objectives, Topologies, Requirements, and Procedures in Service Provisioning of Multi-Cloud

10.2.1.1 Objectives in Service Provisioning of Multi-Cloud

The objectives in service provisioning of multi-cloud includes self-service provisioning, autonomous workload distribution, elasticity, and removal of latency constraint.

Self-Service Provisioning and Autonomous Service Provisioning

The main objective of provisioning in multi-cloud is to provide self-service provisioning which means the user can select the service as per the requirement with less intervention from the cloud service provider (CSP). The service provisioning can also be autonomous which provides service based on user requirements with less user intervention in service selection process.

Autonomous Workload Distribution

The workload distribution which means allocating and releasing resources should be handled in efficient manner in order to avoid infrastructure failures.

Elasticity

The elasticity is also key objective which ensures computational resources in multi-cloud are scaled with flexibility. To ensure elasticity, there are certain approaches to be adopted, namely, load balancing, application scaling, and migration.

Removal of Latency Constraint: Location (Placement and Consolidation)

The latency constraint is another most vital objective which can be removed in multi-cloud environment since the location of data center is provided based on the customer geographical parameters which yield higher availability without interruption.

10.2.1.2 Topologies in Service Provisioning of Multi-Cloud

The topologies associated with service provisioning based on approaches criteria are service provisioning model, brokerage-aided provisioning, policy ensure service level agreements (SLAs), heuristic and holistic perspective, multi-criteria decision-making (MCDM), and algorithmic techniques.

Service Provisioning Model

The Service Measurement Index (SMI) can be used to meet the objective of self-provisioning. This index is a service measurement model which is mainly based on the business model of the International Standard Organization (ISO). SMI model allows the users to choose based on the dimensions provided with various cloud offerings.

Brokerage-Aided Provisioning

The cloud brokers play a prominent role in satisfying the autonomous service provisioning objective. In this brokerage-aided provisioning method, the broker first collects all details of the services of each CSP, analyzes, and creates an index which is utilized when a user request is provided. The cloud performs matching task on the index created and provides the best service to the user.

Policy Ensure Service Level Agreements (SLAs)

Service level agreements (SLAs) are the terms and conditions to which consumer and provider have a mutual agreement. In service provisioning of multi-cloud environment, the process of SLA has to be automated or semi-automated in order to meet the objective of elasticity which in turn reduces the cost and increases the trustworthiness of service provider.

Heuristic and Holistic Perspective

Frameworks and tools have been existing in providing holistic and heuristic task selection and resource allocation options with better resource utilization and energy aware features.

Multi-criteria Decision-Making (MCDM)

This MCDM method is used to select the best service based on the performance measurements provided by the third part monitoring tools. This method is proven to be used for real-world complex problems of service selection.

Algorithmic Techniques

Genetic algorithm and intelligent service provisioning using machine learning techniques have been used as a best search tool for service selection. These techniques have also been used for task scheduling in a dynamic manner and deployment of services in an optimal manner.

Requirements in Service Provisioning of Multi-Cloud

Requirements about multi-cloud is unique for each organization, but patterns of requirements can be identified which is of two broad categories: patterns related to distributed deployment of applications and patterns related to redundant deployment of applications.

Patterns Related to Distributed Deployment of Applications

The concept of this pattern is mainly about the integrations of public clouds from various cloud service providers. Partitioned multi-cloud pattern is a best example under this category. The mechanism of partitioned multi-cloud pattern is to execute an application A in one public cloud of one vendor (Amazon Web Service), whereas application B in another cloud owned by different vendor (Azure). The advantages are avoidance of vendor lock in, and shifting of workloads in different computing environment has been simplified.

Patterns Related to Redundant Deployment of Applications

The concept of this pattern is focused on deployment of replicas of applications in multiple cloud environments in order to increase scalability and availability.

10.2.1.3 Procedures in Service Provisioning of Multi-Cloud

The various procedures involved in the process of service provisioning of multi-cloud are user request scheduling, service selection, service composition, service monitoring, and orchestration.

User Request Scheduling

Cloud services are requested by the cloud consumer which is sent as tasks or jobs at the datacenter of the cloud service provider (CSP). The multi-cloud environment is set up to ensure availability higher than the single cloud environment. Sanjaya et al. [2] have proposed task scheduling algorithm which is allocation aware using traditional Min-Min and Min-Max algorithm for multi-cloud environment.

Service Selection

Service selection is the first and foremost procedure in service provisioning. The user selects the service based on the standards, price, and flexibility. In autonomous service selection process wherein the user intervention is less, the service is selected based on the rank and prediction obtained from the analysis done on the service performance, price, and its features.

Service Composition

Cloud service composition is a procedure of composing distant meta-services which is simpler and satisfies the consumer requirements. Cloud service composition can be categorized as two types, namely, functional and non-functional known as Quality of Service (QoS).

Service Monitoring and Orchestration

Service monitoring is the process of providing measurement information which is required for pricing and also for performance analysis.

Service orchestration is essential for coordinating the process between the services, namely, IaaS, PaaS, and SaaS.

At IaaS level service orchestration in multi-cloud, it requires automation and abstraction to handle different standards of CSP and heterogenous API of each CSP. The automation and abstraction are provided using libraries, standards, and models and cloud orchestration tools. The common libraries used for abstraction are Apache j clouds, Apache Lib cloud, and fog. The standards and models used for automation and abstraction are Open Cloud Computing Interface (OCCI), OASIS TOSCA. The cloud orchestration tools used for abstraction and automation are Apache Brooklyn and Cloudify.

At PaaS level service orchestration in multi-cloud, it focuses on application-centric resources and pre-defined environments. CloudFoundary and OpenShift are used as an API abstracting layer for PaaS.

Cross-level service orchestration is a novel area wherein orchestration is provided across different level of service rather than one service level (IaaS or PaaS). CloudML provides cross-level service orchestration.

10.3 Intelligent Service Provisioning (ISP) in Multi-Cloud

10.3.1 ISP: Methodologies, Advantages, and Limitations in Multi-Cloud Environment

The review on intelligent service provisioning (ISP) in multi-cloud is presented into three categories. The problems and solutions using machine learning techniques in each category has been specified. The three categories incorporated in review are characteristics of service provisioning, approaches of service provisioning, and procedures of service provisioning in multi-cloud. Table 10.1 presents the classification of Intelligent Service Provisioning (ISP).

10.3.1.1 ISP Based on Characteristics of Service Provisioning

The ISP based on characteristics of service provisioning involves workload management, elasticity, and removing latency constraint.

Table 10.1 Classification of ISP

S. no	ISP – classification category	ISP – attributes	Machine learning techniques used
1.	Characteristics of service provisioning	Workload management	Deep belief networks (DBN)
		Elasticity	Reinforcement learning -Q learning
		Removing latency constraint (placement and consolidation)	Support vector machine (SVM)
2.	Approaches of service provisioning	Service provisioning models	1. Non-hierarchical clustering 2. Content-based filtering 3. Behavioral and collaborative filtering 4. Informational filtering
		Brokerage aided provisioning	Multi-learning cloud broker
		Policy ensured service level agreements (SLAs)	Support vector machine (SVM)
		Heuristic and holistic perspective	Reinforcement learning (RL)
		Multi-criteria decision-making (MCDM)	Principal factor analysis (PFA)
		Algorithmic techniques	Deep learning technique – Long short-term memory (LSTM)
3.	Procedures for service provisioning	User request scheduling	Deep reinforcement learning (DRL)
		Service selection procedure	Hierarchical clustering algorithm
		Service composition	Bayesian based model
		Service monitoring and orchestration	1. Random Forest 2. Adaptive boosting classifier 3. Binary logistic regression 4. Neural network 5. Extreme gradient boosting (XgBoost)

Intelligent Service Provisioning for Workload Management

Workload management is the method of managing resources in cloud system. The resources management is of two types, and they are pro-active and reactive management. Reactive management is only providing analysis based on monitored data which is a time-consuming process, whereas proactive management is providing future workload prior for effective management of resources in multi-cloud environment. Intelligent service provisioning (ISP) focusing on workload management is essential for effective resource utilization with Quality of Service (QoS) and less power consumption at the multi-cloud environment. The intelligent workload management process means providing workload prediction and forecasting. In workload prediction and forecasting, it provides the workload of VM in advance which aids VM to auto-scale its resources to fulfill Quality of Service (QoS) and saves consumption of energy. Deep belief network (DBN) is a deep learning technique which has been used in image classification, audio classification, and speech recognition. There are correlations among VMs in multi-cloud environment whose workload can be predicted using deep learning model which consists of DBN and logistic regression. The input given to the deep learning model is the previous CPU utilization, and the top layer which is the logistic regression layer predicts the future workload of all VMs. The output given by the deep learning model is the future CPU utilization of all VMs from which the VMs can auto-scale its resources appropriately. The advantage of workload prediction using deep learning model is that its performance is high since the inherent features of the workload are used rather than using only monitored data [3].

Intelligent Service Provisioning for Providing Elasticity

Elasticity in multi-cloud can enable using auto-scaling mechanisms. Reinforcement learning is one of the auto-scaling mechanism which consists of a decision-making agent which makes decisions based on the experiences and provides the appropriate actions to execute which can be addition and deduction of resources and gaining benefits such as maximum throughput and less response time [4].

Barett et al. [5] incorporated reinforcement learning known as Q learning to provide scaling policies which is optimal in nature, and a Q learning in parallel version has been proposed to reduce the execution time.

Intelligent Service Provisioning for Removing Latency Constraint (Placement and Consolidation)

The placement and consolidation of resources has to be performed efficiently in order to remove the latency constraint. T. Miyazawa et al. [6] provide a way to select resources of non-urgent virtual network for migration automatically and dynamically using support vector machine (SVM) and Q learning with satisfying QoS requirements.

10.3.1.2 ISP Based on Approaches of Service Provisioning

The ISP based on approaches of service provisioning involves service provisioning models, brokerage aided provisioning, policy ensure service level agreements (SLAs), heuristic and holistic perspective, multi-criteria decision-making (MCDM) algorithm, and algorithmic techniques.

ISP Based on Service Provisioning Models

Cloud services selection is a challenging and tedious task for the cloud users especially in a multi-cloud environment. To overcome this challenging task, Rahma et al. [7] have developed a cloud service recommendation system (prototype model) named "USTHB-CLOUD." This model provides recommendations based on content-based and behavior-based analysis from the customer preferences using machine learning techniques, namely, non-hierarchical clustering methods, content-based filtering, behavioral collaborative filtering, and informational filtering. The limitations of this model are that data about service level agreements (SLAs), infrastructure saturation, and energy consumption has not been incorporated in the recommendation system.

ISP Based on Brokerage-Aided Provisioning

Kiran Bala et al. [8] present a machine learning-based broker for decision-making for scheduling the tasks (requests) made by the user at the data center. In multi-learning process of the broker, it learns from all the previous mishandled user requests which are procured as results from single learning. Incorporation of multi-learning method to the cloud broker is made possible through machine learning technique. This machine learning-based cloud broker enhances the accuracy of the decision-making process for each user request for services.

ISP Based on Policy Ensure Service Level Agreements (SLAs)

SLA templates are the special forms of SLAs wherein cloud providers present offers and cloud consumers present requirements before the negotiations for legally signing SLA. SLA mainly consists of three prominent elements, namely, SLA metrics, SLA parameters, and Service Level Objectives (SLOs). They are two types of SLA templates, and they are private SLA which is prescribed for the market customers and public SLA which is formulated for the trade of products in market. C. Redl et al. [9] present an automatic way of matching SLA using machine learning techniques. The intelligent SLA management is possible with proper representation of knowledge from the requirements of the users. Case-based reasoning (CBR) has been used for learning semantically from the requirements. Support vector machine (SVM) has been used as a machine learning technique for providing autonomous SLA matching and autonomous provider selection with cost reduction.

ISP Based on Heuristic and Holistic Perspective

Ali Pahlevan et al. [10] present heuristic and machine learning (ML) approach for provisioning of virtual machines (VMs) into the datacenter. The proposed method in this paper consists of two-level ML approach wherein in the first level of ML, K-means clustering is used for generation of classes and the selection of appropriate classes is performed using heuristic approach. In the second level of ML, value iteration algorithm which is a type of reinforcement learning (RL) is used for classification of classes for provisioning VMs.

ISP Based on Multi-criteria Decision-Making (MCDM)

Analytical Hierarchy Process (AHA) and Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) are the commonly used MCDM methods. AHA provides a comparison of elements in pairwise manner which is structured in hierarchical relationship. Muhammad Umer Wasim et al. [11] have proposed a self-regulated MCDM which used an integrated approach of MCDM and machine learning technique for ranking the service providers. Communality which comes under the category of factor analysis, a type of unsupervised machine learning technique, has been used. Structural Equation Modeling (SEM) has been used for estimation of communality. Commonly used SEM methods are principal factor analysis (PFA) and maximum likelihood. The datasets which had been used for evaluation are customer reviews of cloud service providers (CSP) and simulated dataset of feedback from servers of broker architecture. The benefit of the MCDM with machine learning approach is that it provides ranking based on objective criteria which eliminates error in providing irrelevant services to the customers rather than subjective judgements which is mainly based on insufficient domain knowledge.

ISP Based on Algorithmic Techniques

Service composition is one of the important processes to be performed in multi-cloud environment. Cloud consumers expect the services to be economically feasible. Economically driven service composition is a long process, and it requires exhaustive search. In order to optimize the process of service composition, Samar et al. [12] have presented a deep learning-based service composition (DLSC) framework which is an integration of long short-term network (LSTN) and particle swarm optimization (PSO) algorithm.

10.3.1.3 ISP Based on Procedures of Service Provisioning

The ISP based on procedures of service provisioning involves user request scheduling, service selection procedure, service composition, service monitoring, and orchestration.

ISP Based on User Request Scheduling

Each user request for service to the cloud is considered as a task at the datacenter. Efficient scheduling and service provisioning have to be incorporated at the data center owned by the cloud service provider (CSP) in order to minimize the energy cost at larger scale. Mingxi Cheng et al. [13] proposed a deep reinforcement learning (DRL) approach as a solution to minimize energy cost at the datacenter wherein the user requests for service are scheduled.

ISP Based on Service Selection Procedure

Service selection is the vital procedure in service provisioning. Selecting appropriate services from various vendors in multi-cloud environment is a challenging task. Yan Wang et al. [14] have presented a CC-PSM model which is preference aware service selection model based on customer community. In CC-PSM model, the initial step includes data mining process for categorizing customers based on bipartite network. The second step in this model consists of incorporation of improved hierarchical clustering algorithm which is an unsupervised machine learning technique; the outcome of the second step is to discover consumer community based on preferences. Finally, prediction model is used to predict the customer evaluation on unknown service. The advantage of this model is that it replaces the traditional way of service selection which performs collaborative filtering to find a match between customer requirements and QoS.

ISP Based on Service Composition

Cloud service composition is long term and economic driven. Zhen et al. [15] have proposed a Bayesian-based model to represent the economic perspective of consumer. A novel influence diagram-based model has been presented as a cloud service composition methodology.

ISP Based on Service Monitoring and Orchestration

Service monitoring and orchestration have been achieved using Key Performance Indicators (KPIs). The correlation between resources usage and application of Key Performance Indicators (KPIs) is required for the operation engineer to understand performance bottlenecks, scalability, and degradation of QoS issues. Johannes Grohmann et al. [16] have presented a monitorless model which has a machine learning model trained with platform-related data retrieved from the containerized services to infer KPI. The inferred KPI can be used by the operation engineers to improve the architectural frameworks of multi-cloud in order to fulfill Service Level Objectives (SLOs). The labels used are saturated and non-saturated levels. The machine learning models used in the “monitorless” model are adaptive boosting (AdaBoost) classifier, binary logistic regression, XGBoost, Random Forest (RF), and Neural Network (NN) (Table 10.1).

10.3.2 Comparison of Various Intelligent Middleware for Management of Multi-Cloud Services

A middleware in multi-cloud is defined as a tool which acts as medium between cloud consumer and cloud service provider. A middleware is said to be known as "intelligent" if it has incorporated the computational intelligence techniques, namely, Artificial Intelligence (AI), soft computing, and data mining. Table 10.2 presents the comparison of various intelligent middleware used for management of multi-cloud services.

10.4 Benchmark Case Studies: ISP Provisioning in Multi-Cloud

The two benchmark case studies associated with ISP provisioning in multi-cloud are healthcare and smart city services.

10.4.1 Case Study 1: Multi-Cloud Framework with ISP in Healthcare

Multi-cloud framework has been adopted by healthcare organization to provide agility which is essential to meet the demands of the healthcare sectors for the complete care and support of the patient population. The most popular service healthcare sectors use Software as a Service (SaaS) wherein healthcare acts as a broker to host and maintain applications. Some healthcare sectors have begun to adopt services, namely, Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) in order to deploy cloud native applications.

Ahmed Abdelaziza et al. [30] present a novel model for healthcare services which select the optimal number of virtual machines for processing the medical requests using parallel particle swarm optimization (PPSO), and a prediction model of chronic kidney disease (CKD) is provided using hybrid machine learning techniques, namely, linear regression and Neural Network (NN).

Healthcare requires a large-scale processing of data especially when the data is of image type. Processing large data requires healthcare sectors to adopt multi-cloud systems to provide efficiency, scalability, and high availability. Massive data analysis in multi-cloud of healthcare with more security and less computational costs requires machine learning to be incorporated. Mbarek Marwan et al. [31] proposed a novel approach of enhancing security in multi-cloud of healthcare using machine learning techniques, namely, support vector machines (SVMs) and Fuzzy C-means clustering.

Table 10.2 Comparison of intelligent middleware of multi-cloud services

SI. no	Intelligent middleware	Methodology	Advantage	Disadvantage
1	Cloud management broker (CMB)	Cloud management broker (CMB) is used to manage resources in multi-cloud using two level reinforcement learning (RL) resource allocation algorithm [17]	It provides increased scalability and flexibility in managing resources in multi-cloud with quick response to user requests	The state space size has to be minimized, and scalability algorithms has to be improved using state aggregation mechanisms
2	Daleel – decision-making framework	Daleel is a decision-making framework which provides evidence-based knowledge to customers, and it helps in service selection for customers. Regression-based method has been incorporated in Daleel framework [18]	It provides adaptive decision-making due to machine learning techniques based on response time	The response time of each services of cloud service providers (CSP) has been considered, but other characteristics memory, processor, and behavioral data have not been considered
3	Scalable hierarchical framework	The scalable hierarchical framework is used for resource allocation and power management using deep reinforcement learning (DRL) [19]	It reduces online computational complexity and improves parallelism	Experimental results are provided only with Google cluster traces
4	Brokerage-based cloud service selection	K-nearest neighbor (KNN) is used to search the services of cloud service provider (CSP) index created using B+ tree for providing the cloud service selection [20]	The efficiency of the algorithm is evaluated using real-time and synthetic data	The service level agreements (SLAs) negotiated by the customer at the request time have not been considered
5	Agent-based intelligent cloud service selection	The agent is incorporated with intelligence using unsupervised machine learning technique known as Q learning which provides the appropriate services to the customers through performance checking from the customer feedback [21]	The customer feedback has been considered for service selection which enhances the system to an errorless state	The detailed architecture of the agent has not been explored

(continued)

Table 10.2 (continued)

SI. no	Intelligent middleware	Methodology	Advantage	Disadvantage
6	Trust-based agent learning model for service composition (TALMSC)	TALMSC consists of two-staged fuzzy C-means learning (FCM) mechanism. The efficiency of the mechanism has been tested on JADE which is a multi-agent-based learning system wherein four related mechanism has been tested, namely, two-staged improved FCM, FCM, K- means clustering, and random transaction [22]	TALMSC improves the user satisfaction	The integration of the system with service matching, prediction, and forecasting has yet to be considered and developed
7	Trust enabled self-learning agent model for service matching (TSLAM)	TSLAM is a three-layered agent model which consists of brokers with learning module developed using decision tree algorithm [23]	TSLAM enhances the transaction success rate and improves the user satisfaction	There is saturation level, i.e., the number of cloud service providers to be handled by brokers is limited which has to be increased to enhance efficiency
8	QoS prediction model	A Quality of Service (QoS) prediction model for service composition based on hidden Markov model (HMM) [24]	QoS satisfied service provisioning is provided to the customers through HMM-based QoS prediction model	The prediction model is based on only homogenous services, and heterogenous services were not considered
9	NLUBroker	A reinforcement learning (RL)-based agent is incorporated in natural language understanding (NLU) broker system which maps the customer requirements to the cloud services available from the requirements analysis provided in user language [25]	This NLUBroker provides flexibility to users in providing services requirements	The exploration of natural language processing (NLU) is yet to be done

(continued)

Table 10.2 (continued)

SI. no	Intelligent middleware	Methodology	Advantage	Disadvantage
10	Preference-based cLoud service recommender (PuLSaR)	PuLSar is a cloud service recommender using multi-criteria decision-making approach [26]	It enhances the capabilities of cloud service broker with increased scalability	This recommender has yet to be deployed in real-time cloud environments with exposure of handling heterogenous services in distributed and dynamic environment
11.	Service-oriented broker framework	Service-oriented broker framework consists of three modules, namely, user portal for requirements gathering, service discovery and composition, and service provisioning [27]	The framework improves the main functionalities of cloud	The broker has to be incorporated with machine learning techniques to make it more intuitive in decision-making
12	Cloud broker framework for infrastructure service discovery using semantic network	The broker-based cloud framework provides a user interface wherein the user specifies the request in numerical terms and the broker constructs ontologies and semantic network is formed based on the intersection with the discovered services [28]	The broker-based approach provides a good accuracy in providing service recommendation	The user requirements are specified using numerical representation, and linguistic representation is not accepted
13	Broker-based cloud service model	The broker-based cloud service model executes functions, namely, service discovery and service composition for provisioning [29]	The broker-based cloud service model performs additional services, namely, ranking based on QoS specification	The incorporation of intelligence in broker for federated cloud environment is yet to be developed

10.4.2 Case Study 2: Multi-Cloud Framework with ISP in Industrial IOT and Smart City Services

Basheer Qolomany et al. [32] have proposed an intelligent polynomial time heuristic which selects machine learning models from a superset of model which maximizes the trustworthiness of the model. This selected model is used as prediction model by cloud service providers (CSP) in order to process the data obtained from industrial-based IOT devices and smart city services connected to the cloud.

10.5 Review on State-of-the-Art ML Algorithms for Service Provisioning in Multi-Cloud Environment

Machine learning (ML) techniques come under the category of Artificial Intelligence (AI). Deep learning (DL) is the sub-category of machine learning (ML). Intelligent service provisioning is possible only through the incorporation of machine learning (ML) technique at any procedural level of provisioning or at any service level of provisioning. Machine learning (ML) technique is of two types: supervised machine learning techniques and unsupervised machine learning techniques.

10.5.1 Neural Network (NN)

Working Principle: Neural Network (NN) consists of input layer, hidden layer, and output layer. The Neural Network (NN) is a machine learning algorithm, and its structure is inspired from the biological neurons. It consists of collections of interconnected neurons. The input layer provides input to the hidden layer which computes the output based on activation function, and the final output layer produces the result as the weighted aggregation of all output of hidden layers [33].

Usage: Neural Network ML model is used for multi-cloud service monitoring and orchestration based on Key Performance Indicators (KPIs).

10.5.2 Reinforcement Learning

Working principle: Reinforcement learning (RL) is a type of machine learning technique which is used to control a system to improve its performance stated in numbers. The learner incorporated with RL learns f through the trial-and-error method when exposed to the dynamic environment. There are two approaches used to provide solutions for reinforcement learning problems. The first approach is to search through the dynamic environment wherein genetic algorithms are proved to be effective. The second approach is to use statistical techniques and dynamic programming methods [34].

Usage: The reinforcement learning (RL) ML model is used for multi-cloud user request scheduling in an energy-efficient manner and for provisioning of VMs.

10.5.3 Support Vector Machine (SVM)

Working principle: Support vector machine uses the hyperplane to classify the service instances. The projection of instances to higher-dimensional space occurs if the instances are of non-linear type. Different types of kernels can be used in SVM, namely, Gaussian kernel, polynomial kernel, and radial basis function (RBF) kernel [35].

Usage: The SVM ML model is used for autonomous service selection in multi-cloud based on service level agreement (SLA) attributes.

10.5.4 Deep Belief Network (DBN)

Working principle: The construction of deep belief network (DBN) is in the form of stack wherein each individual unit in stack is built using restricted Boltzmann machine. The structure of restricted Boltzmann machine is a feed forward graph structure with two layers, namely, visible layer which is Gaussian or binary units and hidden layer which is binary unit [36].

Usage: The important aspect to be maintained in multi-cloud is workload management which is provided intelligently using deep belief network (DBN) model.

10.5.5 Principal Factor Analysis (PFA)

Working principle: Principal factor analysis (PFA) is mainly used to reduce the dimensionality of the dataset based on the correlation factor. It mainly focuses on the reduction of non-diagonal elements in the dataset [37].

Usage: The service selection in multi-cloud for user is made at ease by incorporating PFA in the multi-cloud framework which provides user with relevant services based on reviews provided by other users.

10.5.6 Random Forest, AdaBoost and XgBoost

Working principle: Random Forest is an ensemble-based machine learning algorithm. It is a homogenous ensemble classifier since it consists of ensemble with one type of machine learning model, namely, decision trees. It combines the results of many decision trees rather than combining predictions of different machine learning models [38]. Adaptive Boosting (AdaBoost) classifier is a sequential based ensemble classifier, which rectifies error in consequent iterations by giving more weight to the error samples [39]. Extreme Gradient Boosting (XgBoost) classifier is an ensemble classifier which works in parallel and is faster in execution than AdaBoost classifier [40].

Usage: The Random Forest (RF), Adaptive Boosting (AdaBoost),Extreme Gradient Boosting (XgBoost) model is mainly used for service monitoring and orchestration in multi-cloud.

10.5.7 Hierarchical Clustering

Working principle: Hierarchical clustering comes under the unsupervised machine learning algorithm. In hierarchical clustering, the data is grouped into clusters which forms a tree, and the split or merge operations cannot be restructured. There are two types of hierarchical clustering, namely, agglomerative clustering and divisive clustering. The outcome of the hierarchical clustering is in the form of dendrogram which is in the form of tree [41, 42].

Usage: The service selection in multi-cloud based on Quality of Service (QoS) and customer requirements is provided using hierarchical clustering ML model.

10.6 Framework for Intelligent Service Provisioning in Multi-Cloud

The intelligent service provisioning model in the framework consists of process models, namely, service request scheduler, service composition, service monitoring and orchestration, and service selection whose data are analyzed using machine learning (ML)-based decision-making system, and data from cloud service providers, cloud consumers, and process models are stored in knowledge base which is in turn used by decision-making system and process models to satisfy consumer requests.

The framework for intelligent service provisioning in multi-cloud is provided in Fig. 10.1.

The cloud consumer requests are processed by the intelligent service provisioning (ISP) model. The outcomes of process models are as follows:

1. Service request scheduler – sends requests to the corresponding process model
2. Service Composition – The service composition suggestions provided by decision-making system: Services of cloud service providers (CSP) appropriate to user requests
3. Service monitoring and orchestration: The service request status
4. Service selection: Recommendation of services offered by CSP appropriate to user requests provided by the decision-making system

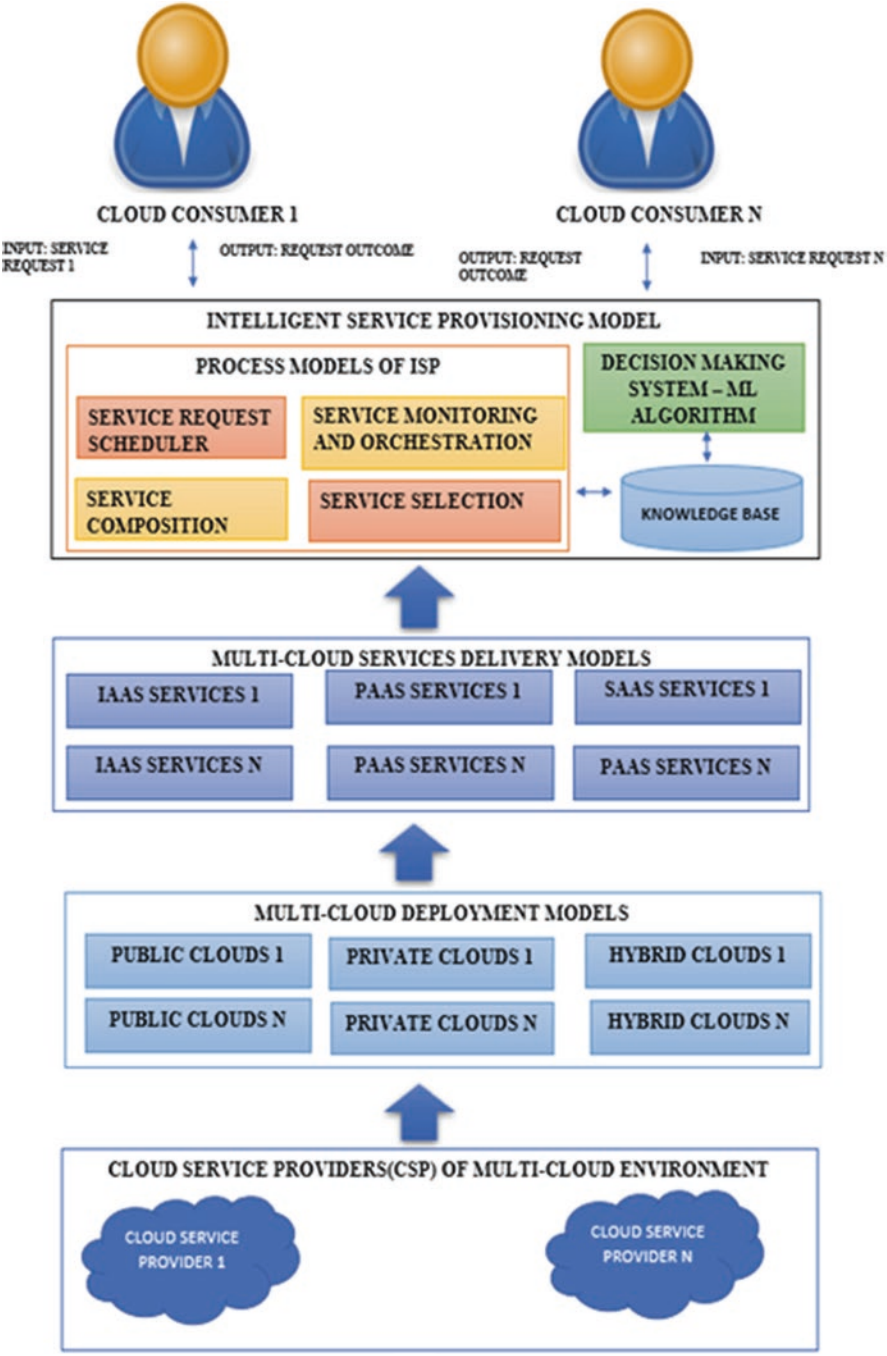

Fig. 10.1 Framework of intelligent service provisioning (ISP)

10.7 Challenges and Future Prospects

The challenges and future prospects of intelligent service provisioning (ISP) in multi-cloud environment have been addressed in this section.

10.7.1 Challenges of Intelligent Service Provisioning (ISP) in Multi-Cloud Environment

1. Provisioning of services with security is a challenge in multi-cloud environment especially sectors associated with finance and healthcare.
2. The costs of services vary, and provisioning of services with optimal cost is a challenging task as it integrates services from various vendor.
3. Service monitoring is a challenging task in multi-cloud environment because the decision has to be made to handle performance degradation.

10.7.2 Future Prospects of Intelligent Service Provisioning (ISP) in Multi-Cloud Environment

1. Integration of block chain with ISP: Security can be enhanced by integrating block chain technology to the ISP framework in multi-cloud environment.
2. Cost: In order to provide optimal cost for services, each services of different vendors has to be registered, and the information has to be stored in knowledge base. The knowledge base has to be dynamic and updated whenever changes in services cost occur.
3. Incorporation of AI: The Deep Learning (DL) algorithm has to be incorporated as decision maker for service monitoring since DL can handle large complex data, and it is faster to provide solutions. DL can be mainly used in healthcare multi-cloud systems wherein complex images have to be stored, processed, analyzed and maintained in cloud.

10.8 Conclusion

Multi-cloud environment provides integrated services from various cloud service providers (CSP). Intelligent service provisioning (ISP) which incorporates machine learning (ML) techniques helps the cloud consumers to procure appropriate services as per the need of the customers. This chapter provides classification of ISP with respect to certain attributes, comparison of various middleware tools used in ISP, and case studied related to ISP.

References

1. Youssef, A. E. (2012). Exploring cloud computing services and applications. *Journal of Emerging Trends in Computing and Information Sciences, 3*(6). ISSN 2079-8407.
2. Panda, S. K., Gupta, I., & Jana, P. K. (2019), Task scheduling algorithms for multi-cloud systems: Allocation-aware approach. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-017-9742-6.
3. Qiu, F. (2016). A deep learning approach for VM workload prediction in the cloud IEEE SNP 2016, Shanghai, China.
4. Dutreilh, X., Kirgizov, S., Melekhova, O., Malenfant, J., Rivierre, N., & Truck, I. (2011). Using reinforcement learning for autonomic resource allocation in clouds: Towards a fully automated workflow. In *7th international conference on autonomic and autonomous systems (ICAS'2011)* (pp. 67–74). IARIA.
5. Barrett, E., Howley, E., & Duggan, J. (2013). Applying reinforcement learning towards automating resource allocation and application scalability in the cloud. *Concurrency and Computation: Practice and Experience, 25*(12), 1656–1674.
6. Miyazawa, T., Kafle, V. P., & Harai, H. (2017). Reinforcement learning based dynamic resource migration for virtual networks. In Proc. IFIP/IEEE symposium on integrated network and service management (IM) (pp. 428–434).
7. Rahma, D. (2017). A novel cloud services recommendation system based on automatic learning techniques. International conference on new trends in computing sciences (ICTCS) https://doi.org/10.1109/ICTCS.2017.58.
8. Bala, K., & Vashist, S. (2014). Machine learning based decision making by brokers in cloud computing. *International Journal of Application or Innovation in Engineering & Management (IJAIEM), 3*(7), 269–273. ISSN 2319-4847.
9. Redl, C., Breskovic, I., Brandic, I., & Dustdar, S. (2012). Automatic SLA matching and provider selection in grid and cloud computing markets. In Proceedings of the 13th ACM/IEEE international conference on grid computing (Grid '12) (pp. 85–94). Beijing, China.
10. Pahlevan, A., Zapater, M., & Atienza, D. (2017). Integrating heuristic and machine-learning methods for efficient virtual machine allocation in data centers. IEEE transactions on computer-aided design of integrated circuits and systems.
11. Wasim, M. U., Ibrahim, A. A. Z. A., Bouvry, P., Limba, T. (2018). Cloud service providers optimized ranking algorithm based on machine learning and multi-criteria decision analysis. https://doi.org/10.20944/preprints201801.0125.v1.
12. Haytamy, S, & Omara, F. (2020). A deep learning based framework for optimizing cloud consumer QoS-based service composition. Springer, Computing. https://doi.org/10.1007/s00607-019-00784-7.
13. Cheng, M. (2018). DRL-cloud: Deep reinforcement learning-based resource provisioning and task scheduling for cloud service providers. IEEE.
14. Wang, Y., Zhou, J.-T., & Tan, H.-Y. (2015). CC-PSM: A preference-aware selection model for cloud service based on consumer community. Hindawi Publishing Corporation Mathematical Problems in Engineering (Vol. 2015, Article ID 170656, 13 p). https://doi.org/10.1155/2015/170656.
15. Ye, Z., Bouguettaya, A., & Zhou, X. (2012). QoS-aware cloud service composition based on economic models. In International conference on service-oriented computing. Springer (pp. 111–126).
16. Grohmann, J., Nicholson, P. K., Iglesias, J. O., & Lugones, D. (2019). Monitorless: Predicting performance degradation in cloud applications with machine learning association for computing machinery. ACM ISBN 978-1-4503-7009-7/19/12. https://doi.org/10.1145/3361525.3361543.
17. Antonio, P., Stefano, B., Francisco, F., Alessandro, G., Guido, O., Martina, P., & Vincenzo, S. (2015). Resource Management in Multi-Cloud Scenarios via reinforcement learning. Proceedings of the 34th Chinese control conference, Hangzhou, China.

18. Samreen, F., Elkhatib, Y., Rowe, M., & Blair, G. S. (2016). Daleel: Simplifying cloud instance selection using machine learning. NOMS 2016–2016 IEEE/IFIP network operations and management symposium, Istanbul (pp. 557–563). https://doi.org/10.1109/NOMS.2016.7502858.
19. Liu, N., Li, Z., Xu, Z., Xu, J., Lin, S., Qiu, Q., Tang, J., & Wang, Y. (2017). A hierarchical framework of cloud resource allocation and power management using deep reinforcement learning. arXiv:1703.04221v2 [cs.DC].
20. Sundareswaran, S., Squicciarini, A., & Lin, D. (2012). A brokerage-based approach for cloud service selection. IEEE fifth international conference on cloud computing.
21. Rabbani, I. M., Muhammad, A., & Martinez Enriquez, A. M. (2013). Intelligent cloud service selection using agents. In P. Meesad, H. Unger, & S. Boonkrong (Eds.), *The 9th international conference on computing and information technology (IC2IT2013). Advances in intelligent systems and computing* (Vol. 209). Springer. https://doi.org/10.1007/978-3-642-37371-8_14
22. Li, W., Cao, J., Hu, K., Xu, J., & Buyya, R. (2019). A trust-based agent learning model for service composition in Mobile cloud computing environments, Special section on mobile service computing with internet of things. https://doi.org/10.1109/ACCESS.2019.2904081.
23. Li, W. , Cao, J., Hu, K., Xu, J., & Buyya, R. (2019). TSLAM: A trust-enabled self-learning agent model for service matching in the cloud market. *ACM Transactions on Autonomous and Adaptive Systems,13*(4).
24. Wu, Q., Zhang, M., Zheng, R., Lou, Y., & Wei, W. (2013). A QoS-satisfied prediction model for cloud-service composition based on a hidden markov model. *Hindawi Publishing Corporation, Mathematical Problems in Engineering*. https://doi.org/10.1155/2013/387083.
25. Xu, L., Iyengar, A., & Shi, W. (2019). NLUBroker: A flexible and responsive broker for cloud-based natural language understanding services. HotCloud'19: Proceedings of the 11th USENIX conference on hot topics in cloud computing.
26. Patiniotakis, I., Verginadis, Y. & Mentzas, G.(2015), PuLSaR: Preference-based cloud service selection for cloud service brokers. *Journal of Internet Services and Applications*. https://doi.org/10.1186/s13174-015-0042-4.
27. Nagarajan, R., & Thirunavukarasu, R. (2020). Service-oriented broker for effective provisioning of cloud services–a survey. *International Journal of Computing and Digital Systems, 10*, 2–17.
28. Nagarajan, R., & Thirunavukarasu, R. (2017). A cloud broker framework for infrastructure service discovery using semantic network. *International Journal of Intelligent Engineering and Systems, 11*(3). https://doi.org/10.22266/ijies2018.0630.02.
29. Rajganesh, N., & Thirunavukarasu, R. (2016). A review on broker based cloud service model, CIT. *Journal of Computing and Information Technology, 24*(3), 283–292.
30. Abdelaziz, A., Elhoseny, M., Salama, A. S., & Riad, A. M. (2018). A machine learning model for improving healthcare services on cloud computing environment. *Measurement, 119*, 117–128. https://doi.org/10.1016/j.measurement.2018.01.022
31. Marwan, M., Kartit, A., & Ouahmane, H. (2018). Security enhancement in healthcare cloud using machine learning. *Procedia Computer Science*.
32. Qolomany, B., Mohammed, I., Al-Fuqaha, A., Guizan, M, & Qadir, J. (2020). Trust-based cloud machine learning model selection for industrial IoT and smart city services. arXiv:2008.05042v1 [cs.LG].
33. Trivedi, M., & Somani, H. (2016). A survey on resource provisioning using machine learning in cloud computing. IJEDR. ISSN: 2321-9939.
34. Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*.
35. Smola, A. J., & Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*, 199–222.
36. Khan, A., et al. (2016). Deep belief networks. https://doi.org/10.13140/RG.2.2.17217.15200.
37. Chapter 420 - factor analysis, ncss statistical software. NCSS.com.
38. Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research*.

39. Ma, C., Chen, F., Liu, J., & Duan, J. (2014). A new method of cloud detection based on cascaded AdaBoost. 8th international symposium of the digital earth (ISDE8) IOP publishing. IOP conf. series: Earth and environmental science. https://doi.org/10.1088/1755-1315/18/1/012026.
40. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. ACM. ISBN 978-14503-4232-2/16/08. https://doi.org/10.1145/2939672.2939785.
41. Rani, Y., & Rohi, H. (2013). A study of hierarchical clustering algorithm. *International Journal of Information and Computation Technology, 3*(11), 1225–1232. ISSN 0974-2239.
42. Pawar, P. S., Sajjad, A., Dimitrakos, T., & Chadwick, D. W. (2015). Security-as-a-service in multi-cloud and federated cloud environments. In *IFIP international conference on trust management* (pp. 251–261). Springer.