

International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN

ENGINEERING

ISSN:2147-6799

www.ijisae.org

Integrating Cloud Services for Comprehensive Cloud Prediction via NWP, LSTM, and K-Means

¹Aiswarya Nekkanti, ²Deva Nandini Kambhampati, ³Asma Pathan, ⁴Dr. Amarendra K., ⁵Rishitha Musunuru, ⁶V. Murali Mohan

Submitted: 02/11/2023 Revised: 21/12/2023 Accepted: 03/01/2024

Abstract: Cloud forecasting, often termed as cloud capacity prediction or cloud resource projection, involves the anticipation and estimation of forthcoming requirements and usage patterns for cloud computing resources. establishing computing assets, such as storage, processing power, and networking, over the internet is what cloud computing means. The ultimate objective of cloud forecasting is to assist businesses in effectively organizing and distributing These resources enable them to meet their business demands while lowering expenditures. The methodology utilized in traditional ways are the foundation for this article, and a new strategy was developed by combining several previously established procedures.

To undertake complex statistical computation and modelling for predictions, cloud forecasting makes use of the capacity and adaptability of cloud computing. Collecting information, preliminary processing, feature design, and the development of models are some of the steps that are engaged. The simultaneous execution and spread of the model-training process made possible by cloud computing makes it possible to analyse enormous datasets more quickly and effectively. In addition, it enables simple scalability because additional processing resources can be added or eliminated as required.

Historical data is acquired for cloud forecasting from a wide range of sources, including databases, sensors, and methods of communication. Then, this data is filtered to eradicate outliers and other objectionable information and change it into a format that permits being looked over.

Index terms- Forecasting, cloud computing, ML algorithms, CloudCast.

1. Introduction

This article is an extension of one we previously published for a conference and the title is Comparative Study on Forecasting Techniques by using Cloud Services. In the above mentioned paper it was based on the Techniques and methodologies that were contributed in achieving the forecasting technique in an optimized way.

The algorithm and approach utilized in this work were created based on our research to obtain precise forecasting information. In terms of cloud forecasting, it mainly refers to a technique for obtaining data from a variety of cloud services in an accurate manner.

In this process, a cloud data set technique was utilized, and there, the acquired information was preserved in the form of graphical representations, etc. Most of these essential elements or features, including data collecting, analysis, model building, capacity planning, scenario analysis and optimization, continuous monitoring and modification, cost management, and cloud provider tools, to make any approach flawless.

Collecting relevant information on the usage statistics and demand for resources of cloud-based services and apps is necessary to undertake cloud forecasting. Metrics including CPU usage, memory utilization, network traffic, and storage consumption are part of this data.

Employing historical usage data, machine learning algorithms, and statistical methods, organizations can analyse trends, patterns, and seasonality in data. They can determine typical usage patterns and anticipate demand changes through this study.

Utilizing predictive models helps businesses to forecast future resource use with accuracy, which is important for efficient use of resources, planning, and budgeting. A connection relating variables that are dependent and

¹Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation Vaddeswaram 522502, Andhra Pradesh, India aiswarya8979@gmail.com

²Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation Vaddeswaram 522502, Andhra Pradesh, India kdevanandini063@gmail.com

³Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation Vaddeswaram 522502, Andhra Pradesh, India pathanasma162@gmail.com

⁴Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation Vaddeswaram 522502, Andhra Pradesh, India amarendra@kluniversity.in

⁵Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation Vaddeswaram 522502, Andhra Pradesh, India rishithamusunuru8@gmail.com

⁶Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation Vaddeswaram 522502, Andhra Pradesh, India muralimohan.klu@gmail.com

independent is established using the predictive modelling method known as linear regression. When there exists a straight-forward connection between variables and resource use, it can be helpful in straightforward cases.

In more complicated situations, businesses might use neural system models that make use of cutting-edge algorithms to analyse immense quantities of data and spot subtle relationships. For instance, firms may employ regression-based procedures like support vector machine algorithms and Random Forests. These sorts are compliant enough to include a variety of inputs such as variables, unpredictable correlations, and inter-factor interactions.

By identifying instances of either elevated or decreased consumption and making adjustments accordingly to prevent over or insufficient utilization of resources, these could be used to enhance allocation of resources. Predictive models can also assist firms in foreseeing changes in resource needs, enabling pre-emptive planning for upcoming opportunities or obstacles.

In order organizations to adequately plan for varied levels of demand, it is imperative that their plans incorporate a variety of scenarios. Taking consideration of these scenarios is critical for the following reasons:

A. Seasonal variations:

Due to events like holidays or climatic shifts, many businesses encounter seasonal variations in demand. Businesses can properly forecast demand during periods of high demand and modify their cloud resources by factoring these swings into cloud forecasting. This keeps resources from being over- or under-provisioned, guaranteeing efficient operations and happy customers.

B. New product launches:

When a fresh good or product is introduced, demand patterns may shift considerably. Businesses can study how this might influence on the needs for resources and make plans for scalability by taking this circumstances into justification when projecting cloud usage. This guarantees that the cloud architecture can handle increasing demand without any downtime throughout the setup time.

C. Marketing initiatives:

Marketing programs, such as sales, discounts, or advertising campaigns, may cause a temporary spike in demand. By integrating such circumstances in cloud forecasting, companies can more accurately predict the predicted traffic spike and install the resources required to handle it. This avoids issues with performance, website breakdowns, or lost sales chances at times of high demand.

Events can produce unexpected traffic increases on websites, as can outside influences like news coverage, viral internet trends, or the unforeseen success of a product. Businesses can analyze past data to find patterns and forecast upcoming traffic peaks by include these types of situations. Then, they can set up their cloud architecture to guarantee dependable performance and accessibility even during unanticipated surges.

Businesses can manage their cloud resources more effectively by being anticipatory rather than reactive by contemplating these distinct scenarios in cloud forecasting.

By minimizing needless overprovisioning and make certain that means are certainly available to properly address growing demand, it aids in cost optimization. Additionally, companies may preserve their competitive advantage by consistently exceeding client expectations and requests, notwithstanding difficult times.

2. Literature Work

The research emphasises the prominence of cloud forecasting in a variety of fields, including agriculture, forecasting the weather, solar energy projection, and climate modelling. The research also explores the challenges associated with cloud forecasting, covering problems such as time and space fluctuations, complex cloud, and data limitations. This academic article contributes to the advancement of cloud prediction methods and offers insightful viewpoints on the challenges and prospective applications in this field.

The research, studies, and analysis discussed in this paper were done to create, enhance, and investigate various approaches for forecasting and estimating future needs and consumption trends for cloud computing resources the objective of this branch of analysis is to strengthen the precision, performance, and usefulness of cloud forecasting through a countless academic and real-world study initiatives.

We choose this field of study that we intend to concentrate on as a means to develop techniques, address problems, explore programs, embrace novel innovations, benchmark performance, foster innovation, discuss constraints present insights, etc.

An essential aspect of the continuing development and enhancement of cloud computing methods is cloud predicting. It contributes to industry best practices, directs cloud resource management decision-making, and is essential for maximizing the use of cloud resources while assuring cost-effectiveness and performance.

3. Data Extraction

This project, which entails gathering vital facts from various sources, intends to analyse and create accurate forecasting models for foreseeing the needs for cloud computing resources.

You can create accurate predictions about future resource use using the data gathered as the basis for training and evaluating forecasting algorithms. An outline showing the steps taken to harvest data for this paper's cloud forecasting research is provided below:

- Identify Data Sources
- Cloud Provider Metrics
- Historical Data
- Data Cleaning where it is used for reliability and accuracy
- Selecting the Forecasting Methods
- Iterate and Refining of one particular method or combing the forecasting methods inorder to optimize as well as to get data accurately.

So these steps were considered in finding the algorithm and data that works for cloud forecasting.

4. Methodologies

1.ARIMA:

Integrated moving average, sometimes known as ARIMA, is a stint sequences forecasting approach that is frequently utilized in industries, including cloud computing and the financial and economic sectors. ARIMA is a useful tool for predicting cloud resource use and requests because it is built to detect and anticipate trends in period sequence data.

A. Advantages:

- Well-Established
- Interpretability
- Simple Implementation
- Useful for Stationary Data and Statistical Foundation.

B. Limitations:

In some cases, complex nonlinear patterns inherent in some time sequences information could go unnoticed because ARIMA assumes linear correlations between past and future values.

Data from time series must be stationary or can be rendered stationary by differencing. Data loss or noise introduction during differencing may happen when exchanging with substantially non-stationary data.

The "order" of the moving average and autoregressive components restricts its capacity to detect long-term dependencies in the data. Very long-term patterns might not be accurately reflected.

The accuracy of forecasting can be significantly impacted by the selection of the ARIMA's characteristics (p, d, and q). It takes experimentation and subject knowledge to choose the best specifications.

ARIMA can be sensitive to outliers or extreme values, which might lead to inaccurate forecasts if not properly addressed.

2. Long Short-Term Memory (LSTM):

Furthered system knowledge techniques like neural networks, in particular certain architectures like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), are used to anticipate cloud utilization of resources in time series. These networks are made to capture complicated patterns and connections in sequential data while getting beyond some of the drawbacks of conventional approaches like ARIMA. Multiple locking mechanisms that control the flow of information are present in LSTM cells, enabling them to selectively recall and forget information across lengthy sequences.

A. Key features of LSTM include:

Memory Cell, Gating Mechanisms, Avoiding Vanishing Gradient

B. Advantages of LSTM :

- Long-Term Dependencies
- Nonlinear Relationships
- Feature Learning
- Applicability to Multivariate Data.
- Flexibility.

C. Limitations:

Complexity and Overfitting: Overfitting can occur in these models, especially when the dataset is limited or when the hyperparameters are not adequately calibrated.

D. LSTM Architecture:



Fig 1:Lstm

The application of LSTM to cloud forecasting has proved successful in capturing complex usage patterns, adapting to changing resource demands, and providing accurate forecasts that go above the capability of conventional techniques like ARIMA.

3. K-Means Clustering:

Unsupervised clustering and recognition of patterns problems are handled by the machine learning algorithm K-Means Clustering. K-Means can be used in cloud forecasting to comprehend trends in data on cloud resource use, identify different usage clusters, and support datadriven choices even if the problem is non commonly utilized for time sequences forecasting approaches like LSTM or ARIMA. It seeks to divide a dataset into a specified number of assortments, with each statistics point be appropriate to the cluster with the nearest mean (centroid).

Pattern Recognition, Usage Profiling, Anomaly Detection, Resource Allocation Strategies, and Capacity Planning may all be done with K-Means Clustering in the context of cloud forecasting.

4. NWP Model

Numerical weather forecasting (NWP) is a challenging method that simulates and forecasts the movement of the planet's atmosphere based on the most recent weather measurements. The physical and dynamic phenomena that control the behaviour of the atmosphere, such as air movement, temperature variations, moisture content, and more, are simulated using NWP models. With the aid of these simulations, meteorologists can predict weather patterns across short- to medium-term time spans.

NWP models begin with a set of initial circumstances, which comprise observed information about the state of the troposphere at that time. This data offers details about the climate at numerous areas throughout the world, including temperateness, influence, moisture, wind rate, and more.

A. Numerical Integration:

NWP simulations apply mathematical equations, such as those for fluid dynamics, the laws of thermodynamic and the conservation of mass and energy, to describe the basic physical mechanisms of the atmosphere. Computational methods are used to numerically solve these equations. Each grid cell in the model corresponds to a volume of air, and the model splits the environment into a threedimensional grid.

B. Time Stepping:

By calculating the mathematical formulas for individually the grid's cells iteratively, the model moves forward in time. The calculations take into understanding many endeavors, including air flow, heat and moisture exchange, and more. To replicate how the atmosphere changes over time, this phase is repeated for progressively longer periods of time.

C. Boundary Conditions:

The upper atmosphere communicates with the oceans and land surfaces, among other parts of the Earth's system. Boundary conditions that explain these dealings are encompassed into NWP models, assisting in the establishment of realistic simulations. The model produces forecasts for numerous weatherrelated variable quantity, including wind patterns, precipitation and warmth. once the simulation reaches the chosen future time. These predictions offer information on the projected changes in the weather over the forecast period.

D. Steps to Implement Nwp model

- Select a Model
- Data Assimilation
- Grid Setup
- Equation Implementation.
- Numerical Solvers
- Time Integration
- Boundary Conditions
- Validation
- Forecasting

E. Limitations of Nwp model

- Resolution Limitations
- Initialization Errors
- Model Physics Simplifications
- Data Assimilation Challenges
- Chaotic Nature of Atmosphere
- Uncertainty Communication
- Observational Gaps
- Complex Terrain and CoastalEffects
- Limited Representation of Atmospheric Composition

Despite these restrictions, NWP keeps becoming better as computational resources rise, the quality of observational data rises, and model physics and methodologies advance. To account inherent uncertainty and boost forecast reliability, forecasters frequently employ ensemble forecasting (performing many simulations with marginally distinctive initial circumstances). In order to get around some of these issues and boost forecast accuracy, hybrid systems that integrate NWP with algorithms for machine learning are also being studied.

5. Implementation of Algorithm

We discover that the amalgamation of NWP and the LSTM with the exploitation of K-means clustering is optimal for foreseeing the forecasting related atmosphere, the climate, wind, etc. because LSTM, ARIMA, and NWP models all have limits in terms of forecasting.

A. Combination of NWP with LSTM and CloudCast:

CloudCast with NWP is a forecasting method that makes application of the strength of Long Short-Term Memory (LSTM) and Geometric Weather Prediction (NWP) neural networks for precise cloud forecasting. This method combines the benefits of LSTM, a particular neural network architecture made for sequence data, and NWP, a well-known meteorological modelling methodology.

Using mathematical models, numerical weather prediction (NWP) projects the state of the atmosphere in the future based on current data. A possibility for NWP tasks is the LSTM (Long Short-Term Memory) kind of recurrent neural network architecture since it excels at handling sequence data. A machine learning approach called K-means clustering is used to combine comparable data points.

As K-means is additionally a machine learning algorithm for clustering, LSTM offers numerous advantages over ARIMA. K-Means is used to group comparable usage patterns into clusters, and LSTM is then applied to each cluster to perform individual time series projections to construct a hybrid model for cloud forecasting.

The data gathered by satellites and nwp graphs are able to be used in this method to group into categories and form different sets of data, which could decrease the time complexity. As a result, we developed this paper using these two methodologies. This strategy can enhance accuracy in forecasting by capturing various consumption patterns across clusters.

B. Step-by-Step Approach and algorithm:

- Data Collection and Preprocessing
- K-Means Clustering
- Cluster Assignment
- LSTM Model Setup & Training
- Cloud Cluster Assignment
- Fusion of Predictions
- Cluster-Based Forecasting
- Final Cloud Forecast
- Combining Forecasts
- Deployment and Monitoring
- Validation and Evaluation
- Continuous Improvement

To forecast meteorological variables for next time intervals, run NWP simulations.

Identify cloud-related characteristics in the NWP forecasts.

Each cloud-related feature based on the NWP should be assigned to a cluster using the K-Means model.

In this step, the precise cloud behavior patterns discovered by K-Means are connected to NWP forecasts.

So the errors that are occurred can be spotted by utilizing the following formulae:

- Mean Squared Error = $\Sigma(y_i f(x_i))^2$
- Root mean squared error = sqrt(mean squared error)
- Mean absolute errors= $\Sigma |y_i f(x_i)|$
- $F = w1 * F1 + w2 * F2 \rightarrow fusion of prediction$

Choose the corresponding LSTM model or composition of LSTM models that have mastered cloud patterns resembling the assigned cluster based on the cluster assignments.

Create cloud forecasts for the time spans relating to the NWP predictions using the chosen LSTM model(s).converge the cloud forecasts from the model(s) based on LSTM and NWP.

The accuracy and thoroughness of cloud predictions are improved by this fusion, which occupies benefit of the advantages of both methodologies.

Preprocessing of the data: NWP data, which contains details like high temperature, compression, moisture, wind momentum, etc., is typically gathered at multiple locations and timestamps.

It is obligatory to pre-process and arrange this data into sequences before feeding it to the LSTM model.

C. Selection and Extraction of Features:

Not all NWP variables may be pertinent for a given prediction job.

Create an LSTM architecture for your NWP work using the LSTM model. The LSTM may be trained to recognize temporal patterns using sequences of previous NWP data as input.

Use historical NWP data for the LSTM model's training. Based on the provided sequence of inputs, the model should eventually be able to forecast the future condition of the atmosphere.

K-means clustering: once the LSTM model has been trained. K-means can be used to group these predictions with the actual NWP observations. Each cluster will stand for a collection of related prediction patterns.

Analysis of the clusters is necessary to comprehend the many kinds of recognized weather patterns.

Use the clusters for a variety of purposes, such as enhancing NWP forecasts or comprehending weather variability.

To further guarantee that the clusters represent actual trends, validate them using historical data.

Utilize K-Means clustering to combine comparable usage patterns based on the features that were retrieved.

Using methods like the elbow method or silhouette score, determine the ideal number of clusters (K).

Silhouette coefficients(i) = (b(i) - a(i)) / max(b(i), a(i))

•



Fig 2: Elbow Point

The elbow point in the fig2 corresponds to where a slope can be seen; the collection of items of various categories that were gathered is shown in the second image; the result of employing the elbow with k-means is shown in the third image.



Fig 3: Before clustering



Fig 4: After Clustering

Here, it is obvious from the figure above that the objects selected for k-means clustering using the elbow technique were divided into distinct clusters to group objects that resemble one another into one cluster, and so on, making it simple to identify the data.

Depending on the K-Means clustering outcomes, place each usage pattern in its appropriate cluster. Utilizing the time series data from each cluster, train each LSTM model.

To maintain sufficient convergence and avoid overfitting, See the training/validation loss.

Forecast future resource use for each cluster using the trained LSTM models.

Restore the original cluster designations using the forecasts from the various LSTM models.

This activity diagram allows us to pinpoint exactly what is involved in using LSTM, NWP, Cloud, and k-means.



Fig 5:ActivityDiagram

A. Advantages of this Approach:

• Cluster-Specific Modeling

This approach allows you to model and forecast different usage behaviours within each cluster, potentially improving accuracy for distinct patterns.

• Variability Adaptation:

Cluster-specific resource utilization fluctuations can be accommodated by LSTM models inside clusters, better capturing dynamic shifts.

• Interpretability:

K-Means clustering produces interpretable clusters that make it simpler to comprehend various usage patterns.

Resource Allocation:

Strategies for allocating resources based on each cluster's unique characteristics can be informed by forecasts for individual clusters.

B. Challenges and Considerations:

• Cluster Validity:

The Caliber and applicability of the collected features have a considerable impact on how well K-Means clustering performs. Poor outcomes could be the solution of incorrect clustering.

• Data Quality:

Clear, accurate, and complete historical data are necessary for accurate forecasting. Preprocessing of statistics is essential.

• Interactions Between Clusters:

This method might not stand competent to fully capture patterns involving interactions between clusters.

• Model Maintenance:

Maintaining several LSTM models, each of which corresponds to a cluster, can be difficult and time-consuming.

• Hyperparameter Tuning:

For both K Means and LSTM models, correct hyperparameter tuning is crucial for predicting.

This K-Means clustering, and LSTM model combo be capable of a potent technique to use sequence modelling and pattern recognition in cloud forecasting. To attain the best results, it calls for rigorous application, experimentation, and domain knowledge.

6. Conclusion and Future Work

Compared to conventional techniques, such as using simply historical data or NWP data, this method may be utilized to predict precipitation and solar radiation with a better level of accuracy. The technique can account for the long-term relationship between climate variables and solar radiation as well as the influence of cloud cover on forecasting thanks to the integration of NWP, LSTM, the cloud, and K-means.

The suggested approach is a broad framework that can be modified to suit particular needs and datasets. It incorporates cloud computing for scalability, LSTM model for temporal pattern identification, NWP preprocessing to incorporate textual input, and K-means clustering for identifying similar weather patterns. The program seeks to improve weather prediction precision and offer insightful data based on past weather patterns.

There are numerous ways to forecast the weather, the environment, the sun, etc. However, we have only employed a few approaches to carry out this process in this study. To reduce the complexity of time and spatial complexity, and even these results in precisely obtaining the information, it is preferable to employ a pattern of several methods as resisted to traditional ways.

Future work may result in additional improvements to this model's design, allowing it to be utilized primarily for forecasting and implementation in the future while also reducing time complexity and accuracy requirements.

Finally, cloud forecasting offers a reliable and scalable method for predicting future events and trends. It enables businesses to make informed decisions and better plan on the footing of accurate and reliable forecasts.

References

- [1] R. W. Saunders, T. A. Blackmore, B. Candy, P. N. Francis and T. J. Hewison, "Monitoring Satellite Radiance Biases Using NWP Models," in IEEE Transactions on Geoscience and Remote Sensing, vol. 51,2013.
- [2] G. Mulder, F. J. van Leijen, J. Barkmeijer, S. de Haan and R. F. Hanssen, "Estimating Single-Epoch Integrated Atmospheric Refractivity From InSAR for Assimilation in Numerical Weather Models," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60,2022.
- [3] Y. V. Wang et al., "Relative Importance of Radar Variables for Nowcasting Heavy Rainfall: A Machine Learning Approach," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61,2023.
- [4] Y. -C. Jin, Q. Cao, K. -N. Wang, Y. Zhou, Y. -P. Cao and X. -Y. Wang, "Prediction of COVID-19 Data Using Improved ARIMA-LSTM Hybrid Forecast Models," in IEEE Access, vol. 11,2023.
- [5] J. Nasir, M. Aamir, Z. U. Haq, S. Khan, M. Y. Amin and M. Naeem, "A New Approach for Forecasting Crude Oil Prices Based on Stochastic and

Deterministic Influences of LMD Using ARIMA and LSTM Models," in IEEE Access, vol. 11,2023.

- [6] W. Safat, S. Asghar and S. A. Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques," in IEEE Access, vol. 9,2021.
- [7] E. Caro, J. Juan and S. Nouhitehrani, "Optimal Selection of Weather Stations for Electric Load Forecasting," in IEEE Access, vol. 11, pp. 42981-42990, 2023, doi: 10.1109/ACCESS.2023.3270933.
- [8] A. Nekkanti, A. Pathan, R. Musunuru, K. Amarendra and D. N. Kambhampati, "Comparative Study on Forecasting Techniques by using Cloud Services," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023.
- [9] G. Roussel, M. Turpin, Y. Bardout and G. Jubelin, "Adaptation of Short-Term Cloud Forecasting to Global Scale and Evaluation of its Contribution to Earth Observing Satellite Planning," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14,2021.
- [10] A. H. Nielsen, A. Iosifidis and H. Karstoft, "CloudCast: A Satellite-Based Dataset and Baseline for Forecasting Clouds," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14,2021.
- [11] W. Dong, H. Sun, Z. Li, J. Zhang and H. Yang, "Short-Term Wind-Speed Forecasting Based on Multiscale Mathematical Morphological Decomposition, K-Means Clustering, and Stacked Denoising Autoencoders," in IEEE Access, vol. 8,2020.
- [12] A. H. Nielsen, A. Iosifidis and H. Karstoft, "CloudCast: A Satellite-Based Dataset and Baseline for Forecasting Clouds," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14,2021.
- [13] A. Zaboli, V. -N. Tuyet-Doan, Y. -H. Kim, J. Hong and W. Su, "An LSTM-SAE-Based Behind-the-Meter Load Forecasting Method," in IEEE Access, vol. 11,2023.