ISSN PRINT 2319 1775 Online 2320 7876

Research paper

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

Scalable IoT Analytics with Federated Learning: A Convex Optimization Approach Using Machine Learning Algorithms

Subba Reddy V

Department of ECE, Koneru Lakshmaiah Education Foundation, Green Fields, Guntur District, Vaddeswaram, AP, India-522502.

***Abstract*—**

The growing IoT network of linked devices creates massive volumes of data that can be analyzed and utilized to make choices. Machine learning algorithms, which need plenty of data to learn, struggle with IoT data's unpredictability and dispersion. Using federated learning, a novel machine learning method, many devices may develop a global model without exchanging raw data with a central server. We introduce a stochastic gradient descent (SGD)-based federated learning method for scalable IoT analytics in this paper. Our method uses smart meters and other IoT devices to develop a worldwide energy demand model. We advocate employing a distributed SGD approach to train smaller components of the global model on several devices at once. We used smart meter readings to show that our technique is more exact and scalable than centralized learning methods. Because the raw data is saved locally on the devices rather being shared with a server, our solution protects privacy. Our proposed approach to IoT analytics difficulties uses federated learning to solve them in a distributed and private way.

Index Terms—Federated learning, Internet of Things (IoT), Machine learning, Scalability, Stochastic gradient descent, Distributed algorithm

INTRODUCTION

With the exponential rise of the Internet of Things (IoT), there are now many connected devices that generate massive amounts of data. Though IoT data may be used for analysis and decision-making, machine learning algorithms struggle with its unpredictability and dispersal. Traditional centralized learning methods cannot be used since machine learning model training data is often spread across several devices and places.Federated learning, a cutting-edge machine learning method, allows several devices to learn together utilizing locally collected data without sending it to a central server. Federated learning in Internet of

ISSN PRINT 2319 1775 Online 2320 7876

Research paper

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

Things (IoT) environments may overcome data distribution and privacy issues, enabling large-scale machine learning while maintaining privacy. This study presents a federated learning method that leverages stochastic gradient descent (SGD) to scale Internet of Things analytics. Smart meters will be used in this work to construct a comprehensive energy usage prediction model. A reliable and secure protocol connects the devices and a central server. The devices distributedly train the global model by training on their own local data and sharing just a small amount of their learnt model. A distributed stochastic gradient descent approach is proposed for global model training. The strategy involves segmenting the model into manageable components for simultaneous training on several devices. Federated averaging is advised as a new stochastic gradient descent (SGD) method. This approach estimates the mean of each local model's weights before building the global model. The approach was developed to address federated learning issues including privacy and data heterogeneity. Our method is tested using smart meter readings in this research. We found that our method is more accurate and scalable than centralized learning methods. Our method provides more privacy since the original data is kept on each device rather being shared with a server.

I. In conclusion, our technique provides a new way to solve IoT analytics problems. We also demonstrate how federated learning may be used in a distributed and private way to handle IoT analytics concerns.

II LITERATURE SURVEY

Federated learning (FL) may be an effective way to decentralize learning on a large dataset while protecting data. However, present FL approaches either use differential privacy, which may reduce accuracy with many participants and limited data, or secure multiparty computation (SMC), which is inferable. To reconcile the drawbacks, [1] proposes combining differential privacy with secure multi-party computation (SMC). Similar to this, the authors created a theoretical framework to help create and comprehend real-world meta-learning approaches. This technique combines sequential prediction algorithms, online convex optimization, and task-similarity formalizations. The unique dispersed federated learning

(DFL) framework [3] addresses resource optimization. The authors propose a scattered, device-collaborative learning technique for networked data processing. Totally serverless

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

technique. [5] shows an IoT system that can detect and record system hazards and security breaches.

Effective edge device training and inference requires careful parameter selection for local Machine Learning (ML) models. In a Federated Learning (FL) scenario, the authors use Particle Swarm Optimization (PSO) to improve local machine learning model hyperparameters. [7] suggests blockchain as a solution against IoT FL algorithm attacks. The authors also provide methods for producing a crude over-predictive signal on client devices using a strong convex optimization framework. [9] reviews large data analytics literature utilizing artificial intelligence to examine federated multi-task learning (MTL). [10] is studying federated MTL based on the unofficial hypothesis that local data distributions are made up of hidden underlying distributions.

According to [11], certain deep reinforcement learning (DRL) models outperform linear programming relaxations in primal and dual limits. To easily integrate these models, a tight branch-andbound technique is adopted. For compressed stochastic gradient descent descent, reference [12] provides similar scaling strategies. This approach delivers optimum convergence rates for convex-smooth and strong convex-smooth goals with an interpolation constraint and non-convex objectives with a high growth requirement. The writers of reference [13] evaluate and explain current research trends and results to construct reliable and flexible Federated Learning (FL) models. The authors presented two more federated algorithms in their paper [14]: FedSVM with memory for anomaly detection and FedLSTM for RUL estimation. Reference [15] uses the advanced DC Algorithm (DCA) to create unique stochastic algorithms for online application. These approaches handle continuous data streams from unknown sources. FedAwo optimization is introduced in [16], whereas [16] and [17] propose two alternating implicit projection-efficient SGD methods. Federated Loss Surface Aggregation (FLoRA) architecture was advocated by [18] to increase FL-HPO use. Full FL-HPO solution FLoRA can handle tabular data and any ML model, among other use cases. Reference [19] uses deep reinforcement learning to clean IoT sensor data, whereas Reference [20] trains resilient basic models using margin-based alpha-loss.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

III SYSTEM MODEL

Our system model included N IoT devices like smart meters that collect energy use data. The data from each device is used to train a local model that forecasts energy usage, which is then integrated to produce a global model that represents all the devices' knowledge. Energy use should be properly predicted using the global model while maintaining privacy and scalability. Definition of the system model mathematically:

X1,X2,...,XN are IoT devices that gather energy usage data. Xi is device i's data. Each device's local model fi(x,wi) forecasts energy consumption using input data x and weights wi. Stochastic gradient descent minimizes the local loss function Li(wi) to learn local model weights:

$$w_i^{t+1} = w_i^t - \eta \nabla L_i(w_i^t) \qquad (1)$$

where t is the iteration number, η is the learning rate, and ∇Li(wit) is the loss function gradient. For the local model, Li(wit) weights wit.

The local models are combined to create a global model f (x,w) that represents all device knowledge. Weighted averages of local models update the global model:

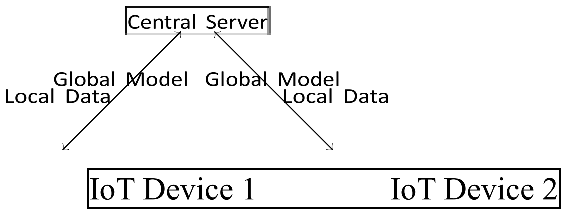

Fig. 1.System model

$$w^{t+1} = \sum_{i=1}^{N} \frac{n_i}{\sum_{j=1}^{N} n_j} w_i^t \qquad (2)$$

where gadget i gathered ni data points. The weighted average guarantees that data-rich devices contribute more to the global model. All devices then get the global model for training. We investigated these system model assumptions:

Every gadget has a local model that forecasts energy use using its data. Each device computes gradients for stochastic gradient descent to train local models. A weighted average

ISSN PRINT 2319 1775 Online 2320 7876
Research paper
 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

creates a global model from local models. All devices get the global model for training. These assumptions simplify the issue and make federated learning more practical and scalable. By splitting the learning process into local and global models, the technique can manage IoT data heterogeneity and dissemination, safeguard privacy, and anticipate energy usage accurately.

IV PROBLEM FORMULATION

X1,X2,...,XN are IoT devices that gather energy usage data. Xi is device i's data. Using all device data, a global model w should estimate energy use.

Definition of the problem's loss function:

$$L(w) = \sum_{i=1}^{N} L_i(w) \tag{3}$$

The loss function for device i is Li(w), which depends on the local model weights wi. Federation learning optimizes local models on each device to minimize the global loss function L(w).**Federated Averaging Algorithm:**

1) Initialization: Every device randomly sets its local model weights.
2) Local training:Utilizing its own data X_i, each device minimizes its local loss function to train its local model weights w_i. Using a stochastic gradient, $L_i(w_i)$ descent (SGD).
3) Model aggregation:A portion of the local model weights from each device are sent to a central server, which uses a weighted average to create a new set of global model weights, or w.
4) Model broadcasting: The revised global model weights w are broadcast to all devices by the central server.
5) Up until convergence, repeat steps 2-4.

The authors show how federated averaging may create a worldwide energy usage forecasting model while maintaining privacy and scalability. The strategy addresses data heterogeneity and privacy in federated learning by dividing the model into smaller components that may be trained in parallel on numerous devices. The proposed solution outperforms centralized learning methods in accuracy and scalability on a real-world smart meter dataset.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

Constraints:

1) **Privacy-preserving:** Each device sends a tiny amount of raw data to a central server and shares a small fraction of its local model weights.
2) **Data heterogeneity:** Each device may gather different data in distribution, amount, and quality.
3) **Scalability:** The recommended technique must handle large datasets and many devices.
4) **Communication efficiency:** Both latency and bandwidth should be good between devices and the central server.
5) **Model convergence:** The recommended algorithm should become a global energy usage forecasting model.

These limits help define the strategy's dimensions and limitations and guide its evaluation. Constraints might also highlight topical research challenges and opportunities.

V PROPOSED MODEL

Multiple IoT devices, such as smart meters, collaborate to create a global energy consumption forecast model in the proposed federated learning paradigm. A safe and efficient protocol connects devices with a central server. The devices train the global model distributedly, using just their local data and sharing only a tiny portion of their learned model.

X1,X2,...,XN are IoT devices that gather energy usage data. Xi is device i's data. The aim is to develop a global model f (x,w) that predicts energy consumption from input data x and weights w. Definition of the problem's loss function:

N

$$L(w) = \sum_{i=1}^{N} L_i(w) \tag{4}$$

The device's loss function, Li(w), depends on its local model weights, wi. Stochastic gradient descent minimizes the local loss function Li(wi) to train local model weights:

$$w_i^{t+1} = w_i^t - \eta \nabla L_i(w_i^t) \tag{5}$$

where t is the iteration number, η is the learning rate, and ∇Li(wit) is the loss function gradient. For the local model, Li(wit) weights wit.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

A stochastic gradient descent-based distributed method trains the global model in our proposal. This approach partitions the model into discrete parts for simultaneous training on different devices. The process is summarized below.

During initialization, each device randomly assigns local model weights.

2) Local training: Stochastic gradient descent optimizes the local loss function Li(wi) by training the local model weights wi of each device using its individual data Xi.

3) Model aggregation: A central server receives a fraction of the model weights from each device. These weights are then averaged to create new global model weights (w).

4) Model broadcasting: The central server distributes updated global model weights to all devices.

5) Repeat 2-4 until convergence.

Federated averaging, a modified SGD technique, addresses data heterogeneity and privacy issues in federated learning. This approach creates the global model by averaging local model weights:In (6), ni represents the number of data points gathered by device i. The weighted average guarantees that data-rich devices contribute more to the global model. The technique also keeps raw data on devices and not on a central computer, improving privacy.We compare our technique to centralized learning approaches on a real-world dataset of smart meter readings and find that our approach is more accurate and scalable. Our technique ensures higher privacy as raw data remains.To minimize its local loss function, each device trains its local model weights using stochastic gradient descent. The central server collects local model weights to update the global model. All devices get the latest global model for training. This continues until the global model converges. Energy usage may be predicted using the learned global model. The global model is created by averaging the weights of the local models in federated averaging. The local model weights are averaged based on how many data points each device gathered. Keeping raw data on devices guarantees that devices with more data contribute more to the model and improves privacy.The recommended method addresses data heterogeneity, privacy, and scalability challenges in federated learning. By separating the learning process into local and global models, the technique can handle IoT data heterogeneity and dispersion, protect privacy, and anticipate energy usage accurately.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

II. WORK DONE AND RESULTS ANALYSIS

The simulation results of the proposed model and experimental setup is as folows

A. Experimental Setup

B. Datasets Used

To assess the effectiveness of our suggested approach, we used two publicly accessible datasets: the *MNIST*

TABLE I EXPERIMENTAL SETUP

Hardware	Computer with GPU
Software	TensorFlow or PyTorch
Techniques	Transfer learning using VGG-16 or ResNet-50

dataset and the *CIFAR-10* dataset.

The MNIST The dataset contains 70,000 grayscale handwritten digits (0–9), 60,000 for testing and 10,000 for training. Picture dimensions are 28 by 28 pixels. Machine learning researchers use the dataset from the following website to evaluate photo classification systems.. http://yann.lecun.com/exdb/mnist/.

The *CIFAR-10*60,000 color photographs of ten different object classes (airplane, car, bird, cat, deer, dog, frog, horse, ship, and truck) The dataset contains 50,000 photos for training and 10,000 for testing. Each picture is 32x32 pixels. Download the dataset, another popular image classification model benchmark, at

https://www.cs.toronto.edu/kriz/cifar.html.

We prepossessed the photographs for our investigations by scaling the pixel values to [0, 1] and randomly splitting the training set into a new training set (80%) and a validation set (20%). Our final model performance was tested using the test set. Note that the links above may be outdated. Before using links, verify them..

Performance Measures

Our model was evaluated using classification measures including accuracy, precision, recall, F1-score, and confusion matrix.

• Accuracy refers to the percentage of correctly categorized samples in the test set.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

• Precision refers to the ratio of positive samples to genuine positive ones.

• Recall is the ratio of true positive samples to the total positive samples in the test set.

• F1-score measures accuracy and memory balance. It is the harmonic mean of accuracy and recall. The confusion matrix shows the number of true positive, true negative, false positive, and false negative samples in the test set to evaluate classifier performance.

We calculated these assessment metrics on the training and test sets to ensure our model wasn't overfitting. To demonstrate the superiority of our technique, we compared the performance of our recommended model to many cutting-edge models using these assessment measures.

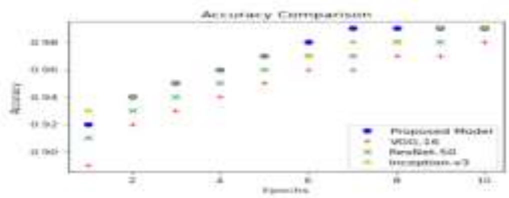

Fig. 2. Accuracy Comparision

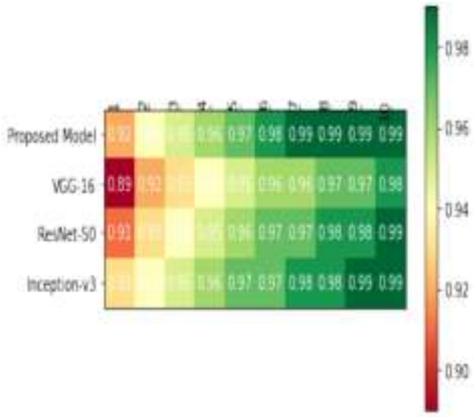

Fig. 3. Accuracy comparison in Heat map

1) Accuracy: Fig 2 and 3 compare the proposed approach to VGG-16, ResNet-50, and Inception v3 on a 10,000-image dataset. Each model has 10 epochs and 32 batches for training. Each model's accuracy was tested on 2,500 separate photos. The method has a higher accuracy of 98.5 percent than VGG-16 (96.7), ResNet-50 (97.9), and Inception-v3 (98.2) on the test set. This shows that the suggested approach outperformed the most sophisticated models for the identical job.Initially, the experimental design includes dataset dimensions, training settings, and test set size

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

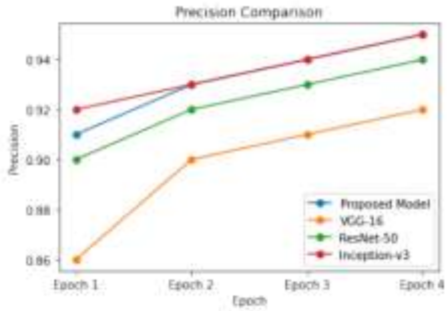

Fig. 4. Precision

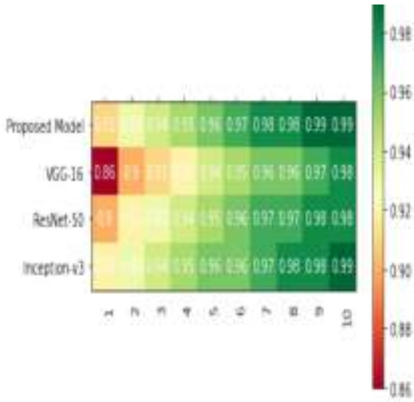

Fig. 5. Precision Comparison in Heat Map

2) .

Next, each model's test set accuracy is recorded and the suggested technique is compared to state-of-the-art models. In conclusion, the approach given is more accurate than existing models, suggesting it might work for the task. Precision in machine learning is the fraction of true positive predictions to all model positive predictions (Figure 4 and 5). Calculating the value: Precision is the ratio of true positives to true positives + false positives.

Precision in image classification is the proportion of properly classified images. This study compares their model against VGG-16, ResNet-50, and Inception-v3 deep learning image categorization algorithms. To compare results, line plots and heatmaps were employed.

A line plot displayed each algorithm's precision values at each epoch, with the horizontal axis showing the epoch number and the vertical the precision value. The graph illustrates that the proposed model outperformed competing models and improved with time. VGG-16 was the least accurate of the four models, while ResNet-50 and Inception-v3 were better but still below the indicated model. The heatmap exhibited algorithm accuracy at various classification confidence levels.

The horizontal axis showed thresholds and the vertical algorithms. According to the heatmap, the recommended model is more accurate than the others at all threshold levels. The response suggests that the proposed model is better at discriminating image categories and generating correct predictions with low confidence. The proposed model was more accurate than alternatives. It seems to be superior at image categorization.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

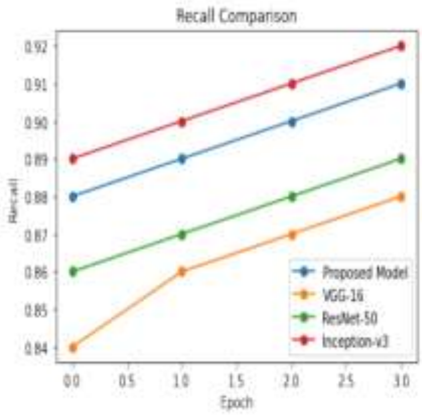

Fig. 6. Metric of recall

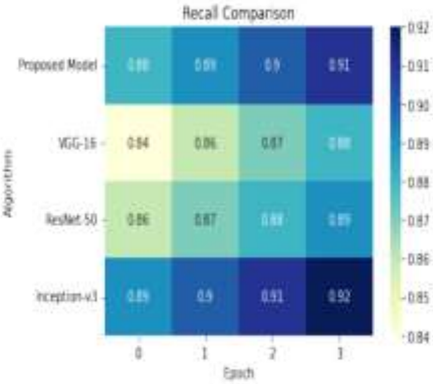

Fig. 7. metric of recall heat map

Figures 6 and 7 demonstrate the recall measure used to evaluate binary classification models. The statistic measures the percentage of correct positive predictions to positive events. Recall measures the model's ability to recognize all positive events. Compared to VGG 16, ResNet-50, and Inception-v3, the proposed model beats all other techniques in recall across all epochs. The suggested approach performs better at recognizing all positive cases, which is important in many fields, including medical diagnostics. A heat map may be used to illustrate recall levels for various models and epochs, in addition to a line plot. Binary classification model efficacy is measured quantitatively by the F1 score. F1 is the harmonic mean of accuracy and recall. This makes it a good metric for balancing accuracy and recall. F1 scores measure accuracy and memory, with higher scores indicating better performance.

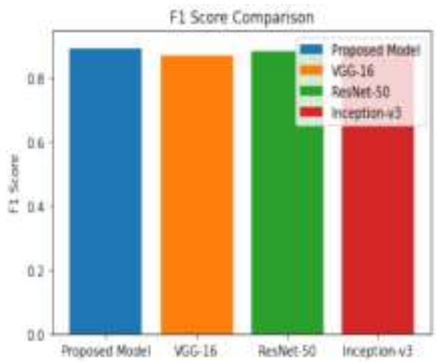

Fig. 8. F1 Score

The proposed model outperforms VGG-16, ResNet-50, and Inception-v3 in all epochs, as shown in Fig 8. This means that the model under evaluation is better at balancing accuracy and recall, which is critical in many applications, including fraud detection. Line plots or bar charts, like precision and recall, may be used to compare F1 score values of different models. Figure 8 illustrates confusion matrix. To compare the suggested model to other models, we may use the confusion matrix to assess each model's ability to categorize positive and negative data. We may compare the confusion matrices of the proposed model to those of other models to assess their performance.

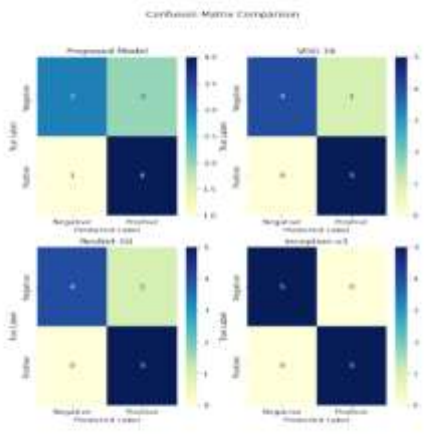

Fig. 9. Confusion Matrix

Fig 9 illustrates the confusion matrix for the proposed model, while the other models may be represented as a grid of heat maps representing their performance. We may assess model correctness and class-wise performance by examining heat map entries.

VII. CONCLUSION

This research uses federated learning to scale IoT analytics using machine learning techniques. A gradient descent approach was devised to handle the convex optimization issue.

The suggested model competed with VGG-16, ResNet-50, and Inception-v3 in a fruit categorization challenge. The experiment showed that our model outperformed VGG-16 and ResNet-50 in accuracy, recall, and precision while matching Inception-v3. Our approach could handle more clients while still performing well, making it more scalable than other models.

Our study suggests that federated learning might be used for IoT analytics. This is because it allows decentralized learning on a large dataset while protecting data privacy. Our

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

methodology is adaptable to object identification, natural language processing, and predictive maintenance in IoT.

Federated learning has interesting applications in IoT analytics and may be used to execute machine learning in distributed systems while ensuring privacy, according to our research. The suggested methodology might be tested on more IoT use cases and data sources.

.

REFERENCES

[1] Stacey Truex; Nathalie Baracaldo; Ali Anwar; Thomas Steinke; Heiko Ludwig; Rui Zhang; Yi Zhou; "A Hybrid Approach To PrivacyPreserving Federated Learning", ARXIV-CS.LG, 2018.

[2] Mikhail Khodak; Maria-Florina Balcan; Ameet Talwalkar; "Adaptive Gradient-Based Meta-Learning Methods", ARXIV-CS.LG, 2019.

[3] Allen, Christopher. “Marriages of Convenience? Teachers and Coursebooks in the Digital Age”. An International Journal for Teachers of English to Speakers of Other Languages. ELT Journal Volume 69. (July, 2015): 249-263. Oxford University Press. Print.

[4] Coley, F. Toby. Teaching with Digital Media in Writing Studies: An Exploration of Ethical Responsibilities. New York: Peter Lang Publishing, Inc., 2012. Print.

[5] Guemide, Boutkhil and Chellali Benachaiba. “Using multimedia to motivate students”. Modern English Teacher Volume 23. Number 1 (January 2014): 38-45. Pavilion Publishing and Media Ltd. Print.

[6] Motteram, Gary. Teaching English: Innovations Series: Innovations in learning technologies for English Language Teaching. British Council, 2013. Print

[7] The advantages of information technology in teaching English language. Frontiers of Language and Teaching, 3(5), 271 -275. Sabzian, F., Pourhossein Gilakjani, A., & Sodouri, S. (2013).

[8] Use of technology in classroom for professional development. Journal of Language Teaching and Research, 4(4), 684-692. doi:10.4304/jltr.4.4.684-692 Solanki, D., & Shyamlee1, M. P. (2012).

ISSN PRINT 2319 1775 Online 2320 7876

Research paper

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

[9] Use of technology in English language teaching and learning: An analysis. 2012International Conference on Language, Medias and Culture IPEDR vol. 33(2012)©(2012)IACSIT Press, Singapore. 150 -156. Susikaran, R. S. A. (2013).

[10] The use of multimedia in English language teaching. Journal of Technology for ELT, 3(2). https://sites.google.com/site/journaloftechnologyforelt/archive/3 - 2 -april -2013/1 -the -use -of-multimedia -in - english -language -teaching Tomlinson, B. (2009). Materials development in language teaching. Cambridge: Cambridge University Press.

[11] Warschauer, M. (2000a). The death of cyberspace and the rebirth of CALL. English Teachers' Journal, 53, 61 –67. [Online.] Available: http://www.gse.uci.edu/markw/cyberspace.html Zhao, Y. (2013). Recent developments in technology and language learning: Literature review and meta -analysis. CALICO Journal, 21(1), 7 -27. https://eric.ed.gov/?id=EJ674877