

MC-SVM Based Work Flow Preparation in Cloud with Named Entity Identification

Dr. Wang Haoxiang,
Director and lead executive faculty member,
GoPerception Laboratory,
NY, USA
Email id: hw496@goperception.com.

Dr. S. Smys,
Professor,
Department of CSE,
RVS Technical Campus,
Coimbatore, India.
Email id: smys375@gmail.com

Abstract: The existing applications that are associated with the internet produce enormous amount of data according to the requirements of diverse circumstances prevailing. This causes multitudes of challenges in examining the data and as well as in the operation of the system that relies on the cloud. To simply process and manage the execution of the tasks properly with respect to time the workflow scheduling was devised in the cloud. To further enhance the process of scheduling the named entity recognition is used. The NER-named entity recognition is an important chore of more general discipline of internet explorer application. Since the NER- problem is highly challenging in cloud paradigm. An innovative frame work termed as the MC-SVM (Multi Class-Support Vector Machine) is laid out in the paper to devise the scheduling of the workflow in the cloud paradigm. The scheduling of the tasks in the cloud delivers an arrangement setting up the work flows with the named entity recognition using the MC-SVM. The algorithm developed enhances the resource allocation process, by performing a simultaneous and dynamic allocation/reallocation of named entities to the resources of the cloud satisfying the demands in the performance and cost. The results observed on validating the proposed algorithm proves the capability of the system to manage the resources in the cloud effectively optimizing the make span and the cost.

Keywords: MC-SVM, Work-Flow Scheduling, NER, Dynamic Allocation, Cloud Computing

1. Introduction

The named entity identification is a significant chore in the process of information extraction. In order to acquire the structured information's using text that are unstructured the named entities are used. So the named entities are defined as the object with the perfect name. The objects may be place, vehicles, people,

facilities, organizations and so on. As the expressions related to time and the numbers are not associated with the entities but very vital in comprehending the text that are unstructured, such details are also incorporated with the named entity recognition.

The work of the named entity recognizer is to identify the entities named and also segregate them. The NER is now broadly engaged in the natural language processing and are used in the POS-tagging, and syntactic chunking. The evolution of the NER began when “Lisa Rau in the year 1992 uses the algorithm to extricate the details of the company names from the financial news. Lisa combined the heuristics with the exception list and the widespread corpus examination to extract the details. Later in the year 1996-2008 the term named-entity for the first time was used in the sixth message understanding conference. As the previous conferences held emphasized only on the “template filling” the sixth conference viewed the sub-tasks that would support the extricating of information’s. This made the “named entities” significant in the year 2003 the author Hammerton applied the LSTM-NN over NER for the German and the English and finds considerable performance improvement for the German than English. The algorithm utilized gathers the information initially and the di-ambiguities and releases the named entities in the later part. Since the percentage of precision and the recall of the NER is lower in ACE 2008 this was termed as an unsolved problem. Meanwhile the common people understood the importance of recognizing the particulars related to a person, place and organizations.

Discovering the sources and recognizing the writing is vital subprojects of the discovering particulars and are termed as the NER. The NER are capable of labelling data automatically with a higher accuracy. The electronic devices should have the knowledge and the ways to recognize a smaller portion of text with the semantic basics to make a proper prediction. To create a perfect association of the tasks in the NLP applications. The paper attempts to recognize the entities using a specific methodology.

A supervised learning technique utilizing the SVM and the enriched features with perception were utilized to distinguish the entities named. Since the MC-SVM are more convincing the SC-SVM in named entity identification in the terms of RBR, LBR, and SB and also reduce the cost and the computation power requirement. So the paper has developed a model using the MC-SVM to schedule the work flow in the cloud with the named entity identification. The general flow diagram presenting the process in scheduling the work flow is shown below in figure.1.

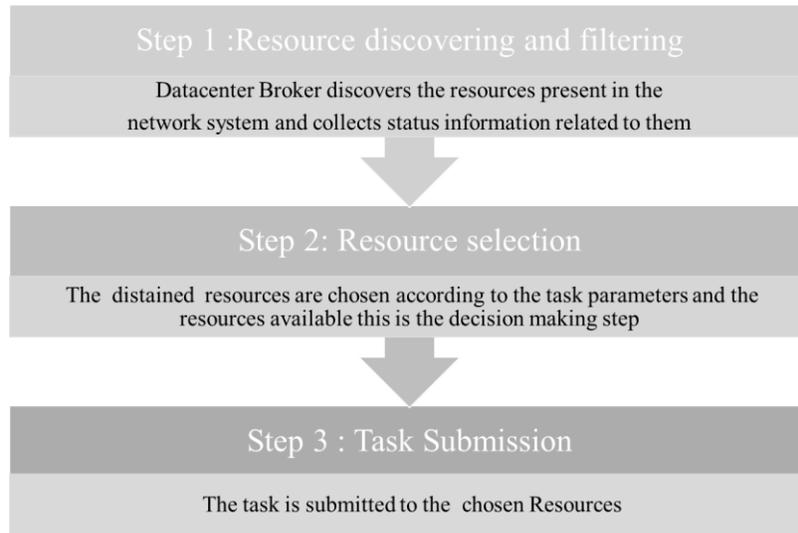


Figure.1 General Work Flow Scheduling Process

The MC-SVM prediction is used in the proposed design to develop the tasks that are data intensive. The algorithm developed enhances the resource allocation process, by performing a simultaneous and dynamic allocation/reallocation of named entities to the resources of the cloud satisfying the demands in the performance and cost. The enhanced resource allocation process laid out is planned with the related works in section 2. Proposed scheduling in section 3. Result Validation using the real time dataset to identify the association as well as the performance of the proposed method in section 4 and conclusion in section 5.

2. Related Works

Liu, et al [1] present the “ cloud work flow system design” elaborating the cloud computing as well as the work flow overview along with the system model depicting the work flow and further detailing the functionality along with the fundamental requirements such as the managing the storage, data duplication, performance and security. Shenai, Sudhir et al [2] focusses on the issues related to preparing the work flows comprehending the various sorts if algorithms used scheduling and delivering the particulars of work flow preparation (WFP) in Grid and Cloud. Simmhan et al [3] conducted the "Building of trident scientific workflow workbench for data management in the cloud."

An broad survey based on the cost effective preparation for work flow in the cloud along with the taxonomy set for the same, analyzing the challenges was performed by Alkhanak et al [4] he deliver the necessary

suggestions for the users and the providers. Kumar, Dinesh et al [5] presents the review on the scheduling the tasks in the ubiquitous cloud. Bhalaji, N. et al [6] proposed a “Delay Diminished Efficient Task Scheduling and Allocation for Heterogeneous Cloud Environment” Raj, Jennifer S et al [7] devised a. "Machine Learning Based Resourceful Clustering with Load Optimization for Wireless Sensor Networks."

Kumar, T. Senthil et al [8] put forth an "Efficient resource allocation and QOS enhancements of IoT with FOG network." Sathesh, et al [9] performed an “Optimized Multi-Objective Routing for Wireless Communication with Load Balancing.” Karunakaran, V. et al [10] utilized hybridized algorithms to devise task preparation for the cloud paradigm Bhalaji, N et al [11] put forth the importance of data duplication in the cloud and proposed an innovative frame work to conduct a secure duplication.

Haixiang, Wang et al [12] utilize the fuzzy petri-nets and the neural networks to identify the optimal routes in the task execution. Bashar, A et al [13] laid out an improved and secure frame work to perform mobile offloading using the cloud paradigm. Candy, Abraham et al [14] framed a resource prediction mechanism for the cloud paradigm for computing enormous data.

3. Proposed Frame Work

The MC-SVM in the preparation of the work flow in the cloud is represented below in the figure.2. The user usually provides the inputs to the NER and mentions the chores to the scheduler. The activities related to the application of the user are performed by the user itself.

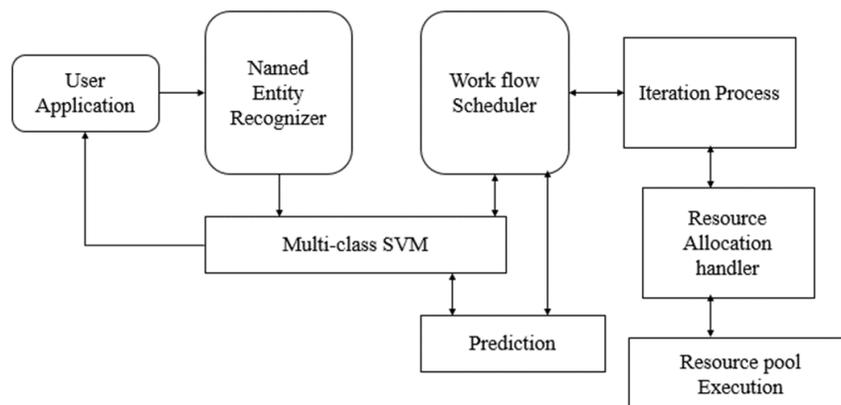


Figure.2 MCSVM Cloud Work Flow Preparation

The named entity recognizer utilizes the qualified casual fields (QCF) to do the distinguishing in the model. The QCF are undirected graph skilled with capability to increase the possibility of the order of the labels provided correspondingly to the order of the input fed. Suppose if the sequence of input is denoted as the $S_i = \{S_{i_1}, S_{i_2} \dots S_{i_n}\}$ and the sequence of labels are denoted as the $S_l = \{S_{l_1}, S_{l_2} \dots S_{l_n}\}$ for the inputs. The possibility of the labels to inputs are calculated using the equation 1

$$prob\left(\frac{label}{inputs}\right) = \frac{1}{z(inputs)} \exp(\sum_{N=1}^n \sum_M \alpha_M f_M(S_{l_{n-1}}, S_{l_n}, S_i, N)) \quad (1)$$

The ‘z’ depicts normalized inputs and f_M denotes the functions of the features, the ‘M’ is the “leaning weight”

The prediction in the MCSVM is done using the “binary classifiers” selecting the prediction and the high confidence rate. The figure. 3 below is the pseudo code for the “binary classifiers”

```

Input : L- learner
Input  $S_i$  and  $S_l$ 
Output classifiers  $f_M$  where  $M = \{1..M\}$ 
begin
{
For every M
Create a new label vector LV= 1
Else
apply L to  $S_i$   $S_l$  to attain the  $f_M$ 
highest confidence rate is achieved
when
LV= arg max prob(input/label)
}
    
```

Figure.3 Pseudo Code for Binary Classifiers

The plane for forecasting gathers the score of confidence (SOC) and generates the outcomes to the work flow preparation on the cloud paradigm. The scheduler links to the process of preparation to execute the applications.

The task scheduler in order to compute more computationally intensive tasks improves the engine with computational procedures that are adaptive and the disseminator disseminates the tasks that are sent by the “work flow engine” to the proper resources chosen by the scheduler. In every step the computation time of the chores, the type of virtual machine and the prominence are delivered back to the device for future examination. The forecasting frame work equips every data set to compute tasks. So the every step is used to prepare the system better.

The device handling the allocation of the resources communicates with the tangible equipment’s that are capable of facilitating the adaptive resource management by identifying the optimal adaptive allocation that are cost effective and less time consuming. According to the behavior of the tasks the system is characterized this improves the allocation process. Scheduler just assigns the tasks and the VM in the resource pool computes the tasks. The algorithm below in the figure.4 describes the enhancement achieved in managing the resources by enhancing the virtual machines and incorporating the tasks

```
Input: Tasks
Output : Resource Allocation
Begin
{
Determine time span
If no time span mentioned
For every task
Do
Sort VM type
Enumerate utility function
For utility function < zero
Enhance the VM to new VM adapting
to task
End if
End for
End for
}
```

Figure.4 Enhanced Resource Management

4. Results Validation

The entity information are assessed according to the monitored and the unmonitored techniques, as well as the set of data used. The method put forth is validated using the real life data sets. These data sets are distinguished and submitted to the scheduler to enhance the process of allocation. The extrication of the association are distinguished into a task and assessed on the basis of precision, recall and the f-score.

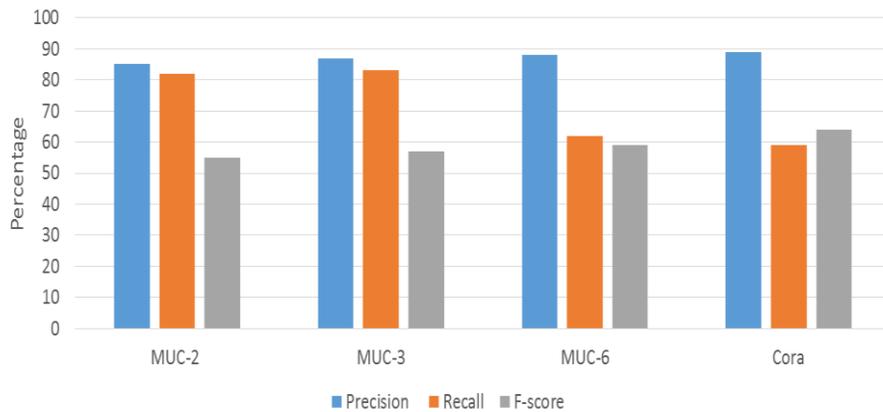


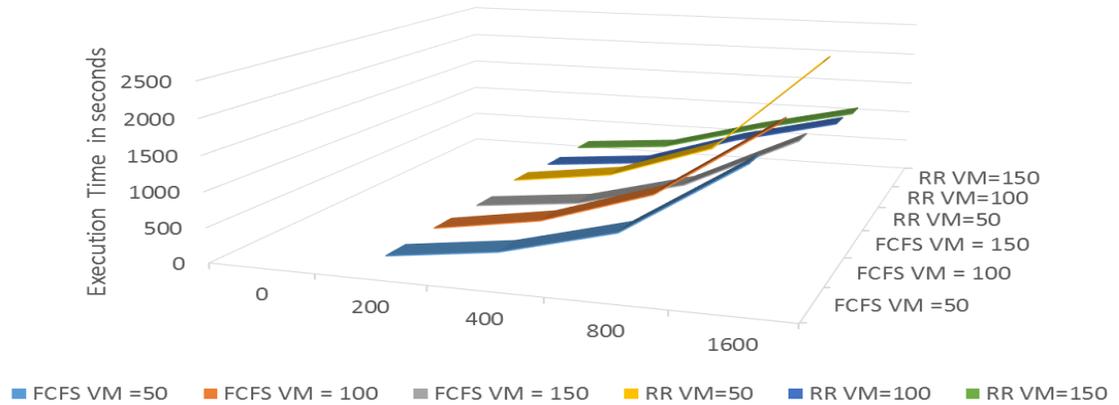
Figure.5 Precision, Recall and F-score for the Data Set

The data set precision, F-score and the recall are displayed in the figure.5 the simulation process for the proposed work is carried out in the work flow-Simulator that is based on Java. The experimental setup and the configuration of the virtual machine is listed in the table.1 below.

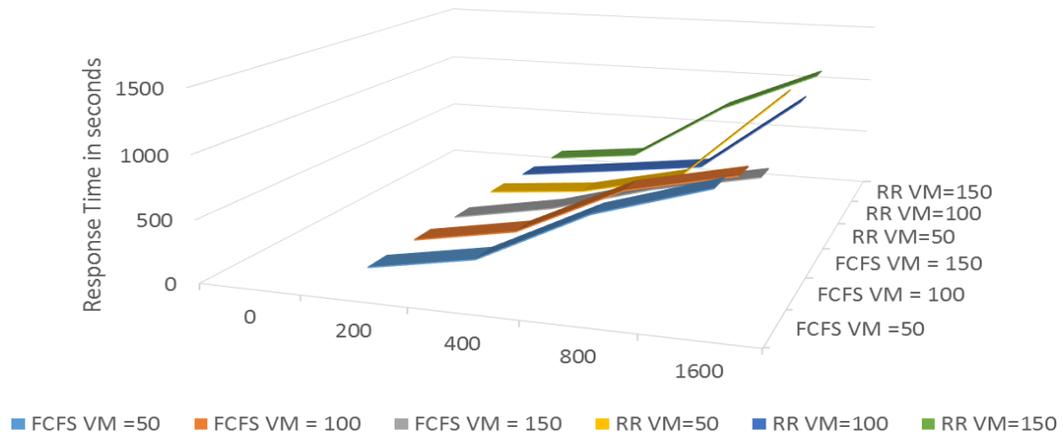
Configuration	Details
Random Access Memory	2 *512
No.of Processors	1
MIPS	250
Bandwidth	1000
Storage Space	10000 MB
Processor type	Ryzen
Scheduling Policy	First Come First Serve , Round Robin

Table .1 Configuration Details

The response time and the execution time for the scheduling based on the first come first serve and the round robin policy is enumerated for the proposed work with varying number of the cloudlets and the virtual machines. The figure .6 (a) and 6(b) is the execution time and the response time achieved using the proposed method for the varying number of virtual machines for the



(a)



(b)

Figure. 6(a) Execution Time 6(b) Response Time first come first serve and the round robin policy

The existing techniques that are engaged in the named entity identification had consumed the major time of execution. "But the work flow preparation with the enhanced resource management proves that the chores provided could be executed in less than a minute.

5. Conclusion

The laid out work in the paper manages a novel problem, the identification of named entity via allocation of resource in the cloud. The MC-SVM method used in the paper is to identify the named entities. The process is integrated with the work flow preparation in the cloud paradigm to have better results. The duration and the cost of demanded are minimized by the proper virtual machine allocation performed using the policies first come first serve and round robin. The Solution obtained indicates that the MCSVM performs well in NER and provides an economical work flow scheduling improving the allocation process by training the model according to the behavior of every chores.

References

- [1] Liu, Xiao, Dong Yuan, Gaofeng Zhang, Wenhao Li, Dahai Cao, Qiang He, Jinjun Chen, and Yun Yang. *The design of cloud workflow systems*. Springer Science & Business Media, 2011.
- [2] Shenai, Sudhir. "Survey on scheduling issues in cloud computing." *Procedia Engineering* 38 (2012): 2881-2888.
- [3] Simmhan, Yogesh, Roger Barga, Catharine van Ingen, Ed Lazowska, and Alex Szalay. "Building the trident scientific workflow workbench for data management in the cloud." In *2009 Third International Conference on Advanced Engineering Computing and Applications in Sciences*, pp. 41-50. IEEE, 2009.
- [4] Alkhanak, Ehab Nabil, Sai Peck Lee, and Saif Ur Rehman Khan. "Cost-aware challenges for workflow scheduling approaches in cloud computing environments: Taxonomy and opportunities." *Future Generation Computer Systems* 50 (2015): 3-21.
- [5] Kumar, Dinesh. "Review on task scheduling in ubiquitous clouds." *J. ISMAC* 1, no. 01 (2019): 72-80.
- [6] Bhalaji, N. (2019). Delay Diminished Efficient Task Scheduling and Allocation for Heterogeneous Cloud Environment. *Journal of trends in Computer Science and Smart technology (TCSST)*, 1(01), 51-62.
- [7] Raj, Jennifer S. "Machine Learning Based Resourceful Clustering With Load Optimization for Wireless Sensor Networks." *Journal of Ubiquitous Computing and Communication Technologies (UCCT)* 2, no. 01 (2020): 29-38.
- [8] Kumar, T. Senthil. "Efficient resource allocation and QOS enhancements of IoT with FOG network." *J ISMAC* 1 (2019): 101-110.

- [9] Sathesh, A. (2019). Optimized Multi-Objective Routing For Wireless Communication With Load Balancing. *Journal of trends in Computer Science and Smart technology (TCSST)*, 1(02), 106-120.
- [10] Karunakaran, V. "A Stochastic Development of Cloud Computing Based Task Scheduling Algorithm." *Journal of Soft Computing Paradigm (JSCP)* 1, no. 01 (2019): 41-48.
- [11] Bhalaji, N. "Efficient and Secure Data Utilization in Mobile Edge Computing By Data Replication." *Journal of ISMAC* 2, no. 01 (2020): 1-12.
- [12] Haoxiang, Wang, and S. Smys. "Soft Computing Strategies for Optimized Route Selection in Wireless Sensor Network." *Journal of Soft Computing Paradigm (JSCP)* 2, no. 01 (2020): 125-137.
- [13] Bashar, A. (2019). Secure and Cost Efficient Implementation of The Mobile Computing Using Offloading Technique. *Journal of Information Technology*, 1(01), 48-57.
- [14] Chandy, Abraham. "Smart Resource Usage Prediction Using Cloud Computing For Massive Data Processing Systems." *Journal of Information Technology* 1, no. 02 (2019): 108-118.
- [15] Chamasemani, Fereshteh Falah, and Yashwant Prasad Singh. "A new boosting multi-class SVM algorithm." *International Journal of Advanced Research in Computer Science* 4, no. 2 (2013).

Authors Biography

Dr. Haoxiang Wang is currently the director and lead executive faculty member of GoPerception Laboratory, NY, USA . His research interests include multimedia information processing, pattern recognition and machine learning, remote sensing image processing and data-driven business intelligence. He has co-authored over 60 journal and conference papers in these fields on journals such as Springer MTAP, Cluster Computing, SIVP; IEEE TII, Communications Magazine; Elsevier Computers & Electrical Engineering, Computers, Environment and Urban Systems, Optik, Sustainable Computing: Informatics and Systems, Journal of Computational Science, Pattern Recognition Letters, Information Sciences, Computers in Industry, Future Generation Computer Systems; Taylor & Francis International Journal of Computers and Applications and conference such as IEEE SMC, ICPR, ICTAI, ICICI, CCIS, ICACI.

Dr. S. Smys received his M.E and Ph.D degrees all in Wireless Communication and Networking from Anna University and Karunya University, India. His main area of research activity is localization and routing architecture in wireless networks. He serves as Associate Editor of Computers and Electrical Engineering (C&EE) Journal, Elsevier and Guest Editor of MONET Journal, Springer. He is served as a reviewer for IET, Springer, Inderscience and Elsevier journals. He has published many research articles in refereed journals and IEEE conferences. He has been the General chair, Session Chair, TPC Chair and Panelist in several conferences. He is member of IEEE and senior member of IACSIT wireless research group. He has been serving as Organizing Chair and Program Chair of several International conferences, and in the Program Committees of several International conferences. Currently