International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 02 Issue: 07 | Oct-2015 www.irjet.net p-ISSN: 2395-0072

SURVEY OF SUPERVISED CLASSIFICATION FRAMEWORK DEVELOPMENT ON HETEROGENEOUS NETWORK

D.Jayachitra[1], Dr.J.Jebamalar Tamilselvi[2]

[1]Research Scholar, Bharathiyar University, Coimbatore, Tamil Nadu, India.
[2]Professor, Department of MCA, Jaya Engineering College, Chennai, Tamil Nadu, India

----------***----------

Abstract

Network Traffic Analysis (NTA) in heterogeneous networks is one of the emerging research areas receiving substantial attention from both the research community and traffic analyzers. Many tasks in NTA can be naturally cast in a supervised and unsupervised learning model. Many supervised classification models and unsupervised clustering learning models in data mining have been proposed for heterogeneous network. Due to the importance of network traffic analysis in data mining research with the rapid development of new models, we provide a comprehensive review on supervised classification on heterogeneous network. In this paper a systematic analysis and comparison of various research works conducted using supervised classification models for network traffic analysis is presented. It addresses the problem of network management such as traffic load, quality of service, and trend analysis. This survey cover real time supervised classification and analyzes techniques for heterogeneous networks. It provides taxonomy of the different supervised classification algorithms and evaluates the various performance metrics that are significantly used for the purpose of comparison. A detailed review is provided covering classification learning algorithms, global voting algorithm and hybrid algorithms. The survey evolve certain open issues, key research challenges for network traffic analysis using supervised classification model in heterogeneous networks, and likely to provide productive research directions.

1. STATE OF ART

The growing population of the aged and the disable is leading to expansion of autonomous service systems. In data mining the data appear in limitless stream for classification of data stream. The problem of data stream classification, where the data enter in an unreal unlimited stream and the probability to evaluate each record is briefed. The problem is solved with the existence of stream classification algorithm.

Bayesian learning and expectation-maximization (EM) techniques were developed under the proposed generative model as shown in [6] for recognizing new training data for learning new unseen sites. Previously unseen attributes combined with their semantic labels were also exposed through another EM- based on the generative model. Besides the space efficiency, the algorithm is time-efficient and highly accurate in [3]. Moreover, one scan algorithm is practical to the heavy hitter problem using distinct elements when compared to the existing fault-tolerant distributed communication techniques.

Anomaly detection aims to recognize a minute group of instances which deviate remarkably from the accessible data. A well-known definition of outlier is that given an observation which deviates so much from other observations, as to arouse the uncertainties behavior generated by different mechanism, it gives the universal idea of an outlier and encourages many anomaly detection methods.

Detecting anomalous insiders in collaborative information systems as shown in [1] intend to analyze the impact of such information in the future. The goal of the current work was to determine the basic information in the access logs and Meta information for the subjects in anomaly detection. On line alert aggregation based on an active, probabilistic model in [7] essentially are regarded as a data stream version of a maximum likelihood approach for the estimation of the model parameters.

Error terms augment the standard sum of squared error computational experiments as shown in [5] that the modified learning method helps to extract fewer rules without increasing individual rule complexity and without decreasing classification accuracy. Ontology-based fuzzy video semantic content model uses spatial/temporal relations in event and conception definitions supply a wide

domain pertinent rule construction average. Fuzzy video semantic content as shown in [4] allows the user to construct ontology for a given domain. In addition to domain ontology additional rule definitions are used to lower spatial relation computation cost and to identify some complex situations more effectively.

Another complexity is that due to isolation requirements and computational problem, it is envisage that classification algorithms are allowed to use only partial information present in the network data and avoid deep packet inspection (DPI). Classification is one of the most frequently encountered decision making tasks. Extending pattern classification hypothesis and design methods to adversarial settings in [8] is extremely pertinent, which has not yet been pursuing in an efficient way.

Data stream classification techniques address the concept-evolution problem which is a major problem with data streams that must be dealt with. A more realistic solution to data stream classification introduces time constraints for postponed data labeling and creating classification decision in [2]. On the other hand, XM properly distinguish among concept drift and concept-evolution, stay away from false detection. Therefore, it fails in considering most of the novel classes as normal data, yielding very high false negative rate.

There are many industrial problems identified as classification problems. For the difficulty of solving such problem precisely lies in the accuracy and distribution of data properties and model ability. Certain points have to be addressed for solving the above said issues. They are mainly:

- Considering seed points in such a way that they are distant enough to be perfectly classified into different categories
- Control the network traffic using supervised classification model to perform non linear dimensionality reduction

2. INTRODUCTION

In this survey paper consider the problem of supervised classification model for network traffic analysis in heterogeneous networks. Figure 1 illustrates the fundamental structure of supervised model.

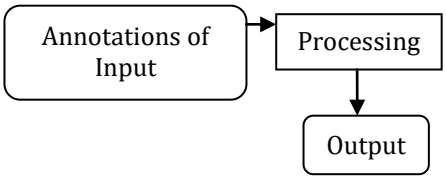

Fig 1 Supervised Model

Supervised can also be combined with the unsupervised model where both input annotations and latent variables are assumed to have caused the output annotations. From the conceptual point of view supervised learning model changes only in the causal structure of the network traffic model. The supervised learning model defines the result of annotations, called inputs, and has another set of annotations, called outputs. In other words the inputs are assumed to be at the beginning step and outputs at the end of the fundamental chain. In supervised learning models, it repeatedly leaves the probability for inputs undefined. The models include intermediate variables between the inputs and outputs.

2.1 Global Voting Algorithm Based on Representativeness

Global Voting Algorithm (GVA) is achieved based on local density and trajectory match information. The sequence of this descriptor over a trajectory gives the voting signal of the trajectory, where high values match to the majority of representative parts. Then, a novel segmentation algorithm is applied on this signal that estimates the number of partitions and the partition borders recognize homogenous division relating to their representativeness. As a final point, a sampling method over the ensuing segments gives up the majority representative subtrajectories in the Moving Object Database (MOD).

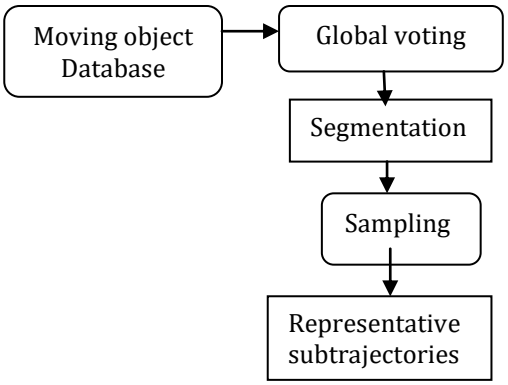

Fig 4 System Architecture of Global Voting Method

GVA is principally a stratified sampling technique, has the restriction that it is user supervised. The superiority of approach is compared to uniform random and stratified sampling techniques. Global Voting Algorithm is described below

Input: An Indexed database

Output: Voting vector V_k

1: For i=1 to L_k

2: $V_k(i) = 0$

3: Repeat

4: Normalized the trajectory voting vector V_k

5: Segmentation of L_k partitions

6: Subtrajectory Sampling Algorithm with normalized lifespan vector

7: Sampling Set Sorted S_k(i)

5: End For

The above steps are used for addressing the issue by segmentation and subtrajectory sampling based on global spatiotemporal similarity of trajectories. GVA extends the density biased sampling from point sets to trajectory segments providing a local trajectory descriptor per line segment that is related to line segment representativeness. Next, Trajectory Segmentation Algorithm (TSA) mechanically and efficiently estimates the number of subtrajectories and their borders, separating each trajectory of MOD into homogenous partitions concerning their representativeness.

To end with, Subtrajectory Sampling Algorithm (SSA) is applied over the resulting partitions providing the most representative subtrajectories of the MOD, also taking into account that high density regions of the MOD should not be oversampled. SSA is terminated by threshold, where the number of moving objects of the original MOD is represented.

2.2 Hybrid approach for context-aware service discovery

An integrated environment intended at providing user's context interest by deploying the semantics entrenched in web services. The main idea of the work is related to augment with qualitative representation of context underlying data by means of Fuzzy Logic in order to recognize the context and to consequently find the right set between the available ones. Semantic formalisms enable the context and services modeling in terms of domain ontology notion. Furthermore, the work defines hybrid architecture which achieves a synergy in the middle of the agent-based example and the fuzzy modeling.

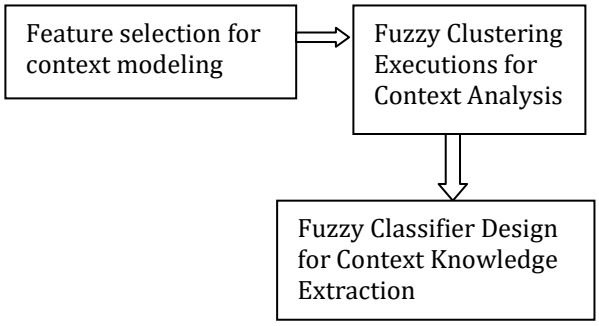

Fig 5 Diagrammatic Form of Context Training Phase

Context Training Phase use techniques of soft computing and semantic web in order to obtain and examine context information. Context Training Phase carries out mathematical models to process context data and trains itself according to the composed knowledge. The process of unsupervised fuzzy data analysis facilitates to augment context modeling with qualitative representation of underling data.

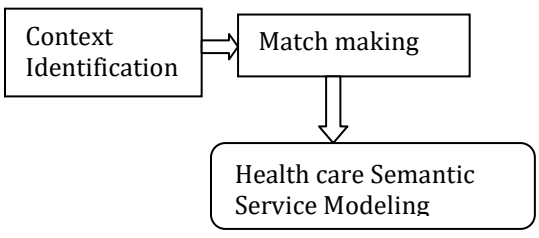

Fig 6 Diagrammatic Form of Context Aware Services Discovery Phase

Context Aware Services Discovery Phase retrieves semantic web services which suitably meet the user's context. A hybrid approach is described based on soft computing and using logic matching evaluation in order to evaluate matchmaking in the middle of parameters and their values. Specifically, the location is stressed when no faithful match occurs between context and services. So, hybrid approach based on soft computing and merely logic matching assessment is defined.

3. COMPARISON OF CLASSIFICATION TECHNIQUE & SUGGESTIONS

In order to compare the execution time of the supervised framework, set of record classes are taken to perform the experiment. The initial metric is the execution time of different existing system, is defined as the time taken to perform the classification process on multi dimensional data. The second performance metric error rate is the number of bit errors occurred on supervised data stream.

The comparison takes place on existing Segmentation and Sampling of Moving Object Trajectories Based on Representativeness via Global Voting Algorithm (GVA), Hybrid approach for context-aware service discovery in healthcare domain (Hybrid Approach). A survey and contribution on classification supervised techniques are available from recent research.

3.1 Measure of Classification Error rate Table

The above table (Table 3.1) describes the error rate of the GVA and hybrid approach. Error percentage of GVA is lesser when compared to the hybrid approach. The raw data illustrating the effects of error rate on different techniques are shown in Fig. 3.1.

More classification supervised techniques developed feature-based and class-based measures searching but the classification accuracy are not improved in diagnosing the network traffic cause. Error rate of GVA is decreased using the partitions with dissimilar degrees of data stream class diversity. Comparatively, it is 75% lesser in GVA when compared with hybrid approach.

3.1 Tabulation for Classification Error rate of different existing technique

Existing Technique	Classification Error rate (%)
GVA	0.4588
Hybrid Approach	0.8512

Fig 3.1 Classification Error rate of different technique

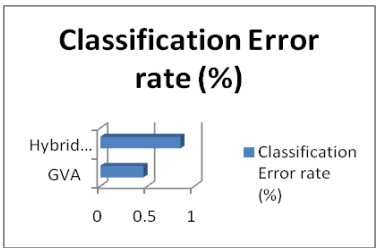

Fig 3.1 demonstrates the error rate of different classification techniques. The usage of clearance in the GVA, a supervised framework decreases the classification error rate when compared with the other existing algorithms. Classification rule improves the searching and classification accuracy reducing error rates on supervised learning. The cause for the network traffic is identified with the class labels in a classifying attributes.

Existing papers has reviewed the potential of the classification learning algorithms. An associative classifier does not follow the classification accuracy maximization paradigm i.e., it commonly portray the training data stream. Survey has reviewed the searching efficiently using classification algorithm for diagnosing the network traffic cause with improved classification accuracy.

Table 3.2: Execution Time

No. of data stream classes	Execution Time (ms)	
	Hybrid approach	GVA
5	702	656
10	794	658
15	717	674
20	716	683
25	724	682
30	726	691
35	738	696

Table 3.2 Tabulation for execution time on supervised framework

As the data stream classes on supervised and unsupervised framework increases, execution time is reduced in the segmentation and sampling of moving objects via GVA. The experiment shows that GVA primarily classifies whole trajectories and greatly brings down the time while performing the execution when compared with the Hybrid approach. Graph shows that the classification via GVA shows conspicuous advantage in controlling the network traffic.

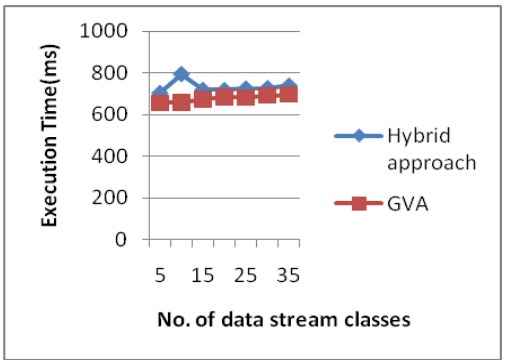

Fig. 3.2 Execution time on supervised framework

Fig 3.2 describes the execution time based on the record data stream classes. The execution time is measured in terms of milliseconds (ms). Classification via GVA is 5 – 20 % lesser delay time taken when compared with the hybrid approach.

The proposal cause a way to forward

- Offers non linear dimensionality reduction to perfectly classify into different categories and control the network traffic.
- Reduced space dimension classifier with integration of supervised and pattern to maintain the heterogeneous traffic.

Supervised learning models on heterogeneous network provides effective network traffic analysis, control and maintenance The cause for the heterogeneous traffic is identified with the class labels data stream in classifying attributes. The presence of the network traffic is identified with the supervised classification algorithm. Based on the training data stream, network traffic control rate is measured. Classification accuracy is enhanced by enhancing the communication level.

5. CONCLUSION

Discussion about the existing Segmentation and Sampling of Moving Object Trajectories via Global Voting Algorithm and Hybrid approach for context-aware service discovery in healthcare domain on execution time and classification error rate parameter dynamically adjusts based on the data stream. Existing Global Voting Algorithm based on local density and trajectory similarity information forms a local trajectory descriptor. GVA primarily clusters whole trajectories and is not customized to recognize the classified patterns of sub trajectories in an unsupervised way. In addition the efficiency is not considered so classification efficiency handling is focused. Hence the classification technique helps in satisfying the efficient way of controlling the network traffic. Extensive experiments evaluate the relative performance of the various algorithms and combinations of supervised model. The result shows that the classification technique performs consistently over a wide range of experimental parameters.

REFERENCES

[1] You Chen., Steve Nyemba., and Bradley Malin., "Detecting Anomalous Insiders in Collaborative Information Systems," IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 9, NO. 3, MAY/JUNE 2012

[2] Mohammad M. Masud., Jing Gao., Latifur Khan., Jiawei Han., and Bhavani Thuraisingham., "Classification and Novel Class Detection in Concept-Drifting Data Streams under Time Constraints," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 6, JUNE 2011

[3] Ying Zhang., Xuemin Lin., Yidong Yuan., Masaru Kitsuregawa., Xiaofang Zhou., and Jeffrey Xu Yu., "Duplicate-Insensitive Order Statistics Computation over Data Streams," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 4, APRIL 2010

[4] Yakup Yildirim., Adnan Yazici., and Turgay Yilmaz., "Automatic Semantic Content Extraction in Videos Using a Fuzzy Ontology and Rule-Based Model," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 1, JANUARY 2013

[5] Thuan Q. Huynh., and James A. Reggia., "Guiding Hidden Layer Representations for Improved Rule Extraction from Neural Networks," IEEE

TRANSACTIONS ON NEURAL NETWORKS, VOL. 22, NO. 2, FEBRUARY 2011

[6] Tak-Lam Wong and Wai Lam., "Learning to Adapt Web Information Extraction Knowledge and Discovering New Attributes via a Bayesian Approach," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 4, APRIL 2010

[7] Alexander Hofmann., Bernhard Sick., "On-Line Intrusion Alert Aggregation with Generative Data Stream Modeling," IEEE TRANSACTIONS ON DEPEDABLE AND SECURE COMPUTING., 2009

[8] Battista Biggio., Giorgio Fumera., Fabio Roli., "Security evaluation of pattern classifiers under attack," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING., 2013