Link Quality Prediction for Multimedia Streaming based on Available Bandwidth and Latency

Lim Su Jin
Univ. Tunku Abdul Rahman
Kuala Lumpur, Malaysia
limsujin@hotmail.my

Lee Sze Wei
Univ. Tunku Abdul Rahman
Kuala Lumpur, Malaysia
leeszewei@utar.edu.my

Simon Lau
Univ. Tunku Abdul Rahman
Kuala Lumpur, Malaysia
simonlau@utar.edu.my

Ettikan Karuppiah
MIMOS Berhad Kuala
Lumpur, Malaysia
ettikan.karuppiah@mimos.my

***Abstract*—Network performance metrics such as available bandwidth and latency are essential to achieve good Quality of Service (QoS) in multimedia streaming. There are unique requirements in network performance metrics for media applications, such as audio conferencing, video streaming, video conferencing, and high-definition (HD) video conferencing. In this paper, we focus on conference call type suggestion based on link quality prediction. The link's quality is classified based on the available bandwidth and latency between two network nodes. We have implemented and compared two of the most popular supervised learning based classification methods, i.e. logistic regression and support vector machine (SVM). We have compared the performance of both methods and their suitability to apply in link quality prediction. The experimental results show that SVM outperforms logistic regression for binary and multiclass classification in terms of accuracy.**

Keywords – multimedia streaming, classification, logistic regression, support vector machine.

I. INTRODUCTION

The on-growing usability in video conferencing system has led to high demand for better quality Audio/Video (AV) transmission. Network constraints are the major problem in multimedia streaming. Network performance metrics serve as optimization parameters for adaptive streaming. Available bandwidth is the maximum throughput between two hosts that can be achieved in the presence of cross traffic while latency is often referred to as time delay between two network nodes. Available bandwidth is important for video conferencing over the Internet to ensure a smooth image quality received. Meanwhile, latency is an essential metric in synchronizing between video and audio received. High latency would cause deteriorated call quality on the receiver and out-of-sync with the video images.

Due to these these considerations and requirements in video conferencing, we propose to predict the link quality between network nodes based on classification of network link based on available bandwidth and latency. Most of the existing approaches only consider either bandwidth or latency solely in their algorithm. Through our proposed supervised learning classification approach, we are able to consider both network performance metrics at once. We suggest to combine both metrics in evaluation based on the requirement of the application. For example, in video conferencing system, available bandwidth and latency plays equally important role in ensuring AV quality in synchronized pattern.

We apply logistic regression and SVM in this paper. These supervised-learning algorithms acquire knowledge from the previous training samples and adapt the system with a new model which is used for prediction on a new input data. Logistic regression (LR) analysis [1] with sigmoid function produces results between 0 to 1. These results can be classified into binomial or multinomial by setting the classification threshold. LR is frequently used to estimate qualitative response models in which the dependent variable is a dichotomy, such as email spam filtering [2], fraudulent detection for online transactions [3], and tumor malignancy classification [4]. The advanced optimization algorithms, such as gradient descent, are often applied in LR analysis to minimize the cost function iteratively.

Support Vector Machine (SVM) [5] is a statistical machine learning technique which learns from a training dataset and attempts to generalize and make correct predictions on new input data. The kernel-based SVMs are able to handle many types of data within the same model which encourage the flexibility of the learning algorithm. SVM has been successfully applied in many applications, such as TCP traffic classification [6], text classification [7] and more commonly in bioinformatics [8].

The key research contributions of this paper are listed as follows:

- Novel link quality prediction via LR and SVM classification are performed.
- Network performance metrics (available bandwidth and latency) are associated in pairs in classification which is useful for interactive AV applications.

The rest of this paper is organized as follows. In Section II, the related works of link quality prediction are discussed. It is followed by the implementation detail of logistic regression and support vector machine in Sections III and IV. The evaluation methodology used in this paper is explained in Section V. The experimental results and discussion are presented and discussed in Section VI and finally concluded in Section VII.

II. RELATED WORK

Network performance prediction is very relevant to the reduction of the overhead associated with continuous measurements. Current approaches for carrying out network performance prediction are based on network latency and available bandwidth normally measured and analyzed

978-1-4799-0539-3/13/$31.00 ©2013 IEEE

separately. Related works for latency prediction include Vivaldi [9], Global Network Positioning (GNP) [10], and Internet Distance Estimation Service (IDES) [11]. Vivaldi faces challenge in violation of triangle inequality while the accuracy of GNP and IDES are influenced by the landmark node selection. The prediction of available bandwidth as proposed by Last-Mile Model [12] is only limited to small scale of data where the accuracy is highly dependent on the number of neighbours selected to iteratively minimize the discrepancies between measured and predicted values. The tree embedding approach, introduced in Sequoia [13] faces constraints on triangle inequality violation, inability to predict asymmetric information, and the influence of the node selection process. Direction-aware embedding is proposed in [14] separating upstream from downstream properties of the hosts to tackle the limitations faced in Sequoia.

Work presented in [15] is the only related work that predicts end-to-end performance classes through decentralized approach based on stochastic gradient descent. The algorithm performs matrix completion with binary performance measures between network nodes with known and unknown nodes to be filled. This approach can be applied on both available bandwidth and latency, but the implementation for latency and available bandwidth is separated. The prediction is also performed in quantitative way, by filling the matrix with measured distance metric and predict the unmeasured through matrix completion (DMFSGD) [16]. The authors have recently extended their work to perform ordinal rating of network performance by network inference based on performance ratings [17].

We have performed linear regression analysis, by optimizing matrix completion problem with stochastic gradient descent in our previous work [18]. The unmeasured distance metrics are first predicted through interpolation, and initialized with singular value decomposition (SVD) before the optimization step. It has proven to be effective in improving the convergence of algorithm to global minimum. This paper is an extension to our previous work, to further enhance the prediction and adapt the algorithm to AV streaming applications through binomial and multinomial classification. The available bandwidth and latency are associated in pairs from our prediction algorithm previously.

LR and SVM classification approaches are described in the following section.

III. LOGISTIC REGRESSION (LR)

Logistic regression [19] is a regression technique suitable for data with binary outcomes {0, 1}. It builds a model from training samples and predicts the probability of the network link to be 0 (Good) or 1 (Bad). The input features are the available bandwidth and latency. In this section, we discuss how the class-based link quality is predicted through LR, whereby a set of training samples will be trained to obtain learning parameter, θ and regularization coefficient, λ for regularized logistic regression through gradient descent.

A. LR Model

Let X be a dataset with dichotomous outcome, $y = \{0, 1\}$. For each training sample x_i in X, the outcome is either $y_i = 1$ or $y_i = 0$. The experiments outcome with $y_i = 1$ are said to have 'Good' link quality, while for $y_i = 0$ for 'Bad' quality.

In supervised-learning, to make sure the input dataset is learnable, a differentiable function is needed to do the fitting instead of using two line segments. The probability that $y = 1$, given x, parameterized by θ or often referred to as logistic regression hypothesis is defined as:

$$p(y=1\,|\,x;\theta) = h_\theta(x) = g(\theta^T x) \tag{1}$$

where function g is the sigmoid function and $h_\theta(x)$ is interpreted as the estimated probability that $y = 1$ on input x. Before executing the actual cost function, a sigmoid function is employed. The sigmoid function is defined as:

$$g(z) = \frac{1}{1+e^{-z}} \tag{2}$$

The training set $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})\}$ with m samples where $x \in [x_1, x_2]^T$ is the set of input features (available bandwidth and latency) used to obtain the fitting parameter, θ to minimize the cost function $J(\theta)$ as follows [19]:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}[-y^{(i)}\log(h_\theta(x^{(i)})) - (1-y^{(i)})\log(1-h_\theta(x^{(i)}))] \tag{3}$$

In order to optimize the algorithm, gradient descent is applied iteratively as follows:

Optimization algorithm:

i. Compute cost function $J(\theta)$.

ii. To $min_\theta J(\theta)$, perform gradient descent:

iii. Repeat {

iv. $\theta_j := \theta_j - \alpha \dfrac{d}{d\,\theta_j} J(\theta)$ (for j = 0, 1, ... n)

v. }

The gradient of the cost function is calculated iteratively to achieve convergence to global minimum. The gradient of the cost is defined as follows:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m} h_\theta((x^{(i)}) - y^{(i)})\, x_j^{(i)} \tag{4}$$

The trained dataset will output a model with decision boundary and trained parameter for new input samples. For the non-linearly separable dataset, features are mapped into higher dimension with higher polynomial terms of x_1 and x_2 to fit it, and regularization term is use to do parameter tuning, θ. The polynomial is expanded up to the sixth power which is best fit in this case. The learning problem can be difficult if the dimension is too high.

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \ldots) \quad (5)$$

Features x_1 and x_2 correspond to the available bandwidth and latency on certain link in the network. The cost function with regularization term is as follows [19]:

$$J(\theta) = -[\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))] + \frac{\lambda}{m}\sum_{j=1}^{n}\theta_j^2 \quad (6)$$

This allows us to build a more expressive classifier which is not susceptible to under-fitting or over-fitting. Gradient descent is applied for optimization as well. The gradient of the cost is defined as follows:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} + \frac{\lambda}{m}\theta_j \quad (7)$$

for j = 1, 2, 3,…, n.

B. Multiclass Logistic Regression (Multiclass LR)

To enable better classification, we implemented four classes by applying "one-against-all" multiclass LR. The hypothesis of multiclass LR is different from that of binary LR while the cost function is similar in both cases. The hypothesis of N-class multiclass logistic regression is:

$$h_\theta^{\,i}(x) = p(y = i \mid x;\theta) \quad (\text{for } i = 1, 2,\ldots,N) \quad (8)$$

where a logistic regression classifier $h_\theta^i(x)$ is trained for each class i to predict the probability that $y = i$. In "one-against-all" multiclass LR, for a new input x, prediction is made by picking the class i that maximizes $max_i\ h_\theta^i(x)$.

IV. SUPPORT VECTOR MACHINE

SVM is suitable for classification problems with high dimensional feature space and small training set size [20]. There are four common kernels in SVM: linear, polynomial, radial basis function (RBF), and sigmoid. In this paper, we focus on C-support vector classification with linear and RBF kernel to study if the distribution is linearly separable or non-linearly separable.

A. *C-Support Vector Classification (C-SVC)*

The main parameter in C-SVC is C, the balance parameter, which plays the role similar to $1/\lambda$, regularization parameter in RLR. The training features with l samples are interpreted as vectors in C-SVC, such that $x_i \in R^n, i = 1,\ldots,l$, and the label vector $y \in R^l$ such that $y_i \in \{0,1\}$, as introduced in [21] to solve the following primal optimization problem.

$$\begin{aligned} &\min_{\omega,b,\xi} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & y_i(\omega^T\phi(x_i) + b) \ge 1 - \xi_i, \\ & \xi i \ge 0, i = 1,\ldots,l, \end{aligned} \quad (9)$$

where the kernel function, $\phi(x_i)$, maps x_i into a higher-dimensional space, $C > 0$ is the regularization parameter, b is a bias, ω is the feature vector and $\sum_{i=1}^{l}\zeta_i$ is the sum of errors in addition to $\omega^T\omega$. Due to the possible high dimensionality of vector variable ω, usually the dual problem is solved as follows:

$$\begin{aligned} &\min_{\alpha} \frac{1}{2}\alpha^T Q\alpha - e^T\alpha \\ \text{subject to} \quad & y^T\alpha = 0, \\ & 0 \le \alpha i \le C, i = 1,\ldots,l, \end{aligned} \quad (10)$$

where e = $[1,\ldots,l]^T$ is the vector of all ones, Q is an 1 by l positive semi-definite matrix, $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, and $K(x_i, x_j) \equiv \phi(x_i)^T\phi(x_j)$ is the kernel function.

After solving the equation in (10), by using the primal-dual relationship, the optimal ω satisfies

$$\omega = \sum_{i=1}^{l} y_i\alpha_i\phi(x_i) \quad (11)$$

and the decision function is

$$\text{sgn}(\omega T\phi(x) + b) = \text{sgn}(\sum_{i=1}^{l} y_i\alpha_i K(x_i, x) + b). \quad (12)$$

Then, we store $y_i, \alpha_i \forall_i$, b, class label names, support vectors, and other information such as kernel parameters in the model for prediction. In this paper, we compare linear and RBF kernel to fit our problem.

Linear kernel: $K(x_i, x_j) = x_i^T x_j$ (13)

RBF kernel: $K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2, \gamma > 0$ (14)

The balance parameter C and kernel parameters are then chosen according to the selected kernel. RBF kernel is able to map samples into higher dimensional space with non-linear attributes, while linear kernel is able to handle linear relation between class labels and attributes only.

TABLE I. BANDWIDTH REQUIREMENT FOR SKYPE[22] AND TRUECONF[23]

Application	Call type	Minimum download speed (kbps)	Minimum upload speed (kbps)
Skype	Calling	30	30
	Video calling/scree sharing	128	128
	Video calling (high-quality)	400	400
	Video calling (HD)	1200	1200
	Group video (3 people)	512	128
TrueConf	Video Call HD	128	128
	Video lecture (speaker)	1920	128
	Video lecture (listener)	128	128
	Virtual meeting (speaker)	384	128
	Virtual meeting (listener)	512	-

TABLE II. THRESHOLD SETTINGS FOR BINARY AND MULTICLASS CLASSIFICATION

Classification	Link Quality	Available bandwidth	Latency	Suggested call type
Binary	Good	≥ 1Mbps	≤ 0.05s	Video calling (high-quality), conferencing
	Bad	< 1Mbps	> 0.05s	Video calling (2 people)
Multiclass	Very Good	≥ 3Mbps	≤ 0.03s	Video calling (HD), conferencing
	Good	≥ 1Mbps	≤ 0.05s	Video calling (high-quality), conferencing
	Moderate	≥ 100kbps	≤ 0.5s	Video calling (2 people)
	Bad	< 100kbps	> 0.5s	Calling (audio)

V. EVALUATION METHODOLOGY

The aim of a learning algorithm is to train a set of training samples to obtain the best fitting parameter. For both classification methods, we evaluate the accuracy as:

$$Accuracy = \frac{P_{correct}}{m} \times 100\% \tag{15}$$

where $P_{correct}$ is the number of correctly predicted data, m is the total number of data. If a trained model has high accuracy on the cross-validated training dataset, then it is assumed to fit to all other new samples, and can be used to classify new samples. The datasets used are from the *bedibe* project [23] and is divided in the ratio of 600:200:200 for training set, cross-validation set and test set.

A. Cross-validation and Grid Search

There is no tuning parameter in LR, while the only parameter λ, regularization parameter in regularized LR, is trained with different values to obtain the best fitting parameter for the algorithm. The learning model is generated from the training set, and the model is validated with cross-validation dataset.

The regularization parameter, C in C-SVC with linear kernel is experimented in the same way, by generating learning model with different parameter. Using linear kernel, large value of C will tend to minimize misclassification which smaller C values would maximize the margin between boundaries.

For C-SVC with RBF kernel, there are two parameters: C and γ. The selection of parameters for C-SVC with RBF kernel is performed through grid search. The different pairs of (C, γ) values are tested and the one with the best cross-validation accuracy is picked. We used 5-fold cross-validation in the experiments to prevent over-fitting problem. We have deployed LIBSVM [22] to run our experiment C-SVC.

B. Classification Threshold

The classification threshold settings are based on the applications requirement, such as bandwidth requirement for Skype [24] and TrueConf [25] as shown in Table I. The incoming bandwidth requirement for video lecture (speaker) in TrueConf is higher to support the continuous media streaming. Interactive communications have stringent requirement in delay. The acceptable values for one-way delay are within 150 to 400ms [26]. Therefore, link with latency larger than 500ms with 5Mbps available bandwidth is predicted as bad. Though with high available bandwidth, the video image received is good but the high latency is causing the audio to be out-of-sync.

In order to secure maximum performance, we have added 50% safety margin to the network performance metrics. With video conferencing being the key interest in our research, the threshold is set accordingly as in Table II. For example, with the 'Very Good' quality, we are able to start a group video with three parties while if the quality is 'Bad', only audio calling is preferable.

VI. RESULTS AND DISCUSSIONS

This section shows experimental results in terms of accuracy for both LR and SVM respectively.

C. Experimental Results for LR

The experimental results for binary classification through LR are shown in Table III.

TABLE III. BINARY CLASSIFICATION BY LOGISTIC REGRESSION

	Training Set	**Cross Validation Set**	**Test Set**
Accuracy (%)	98.17	99.00	99.50

The binary classification obtained through LR presents high accuracy as shown in Table III, with 98.17% of training data are correctly classified. The training set data is plotted as shown in Fig. 1 to illustrate the binary link quality classification. The two axes in Fig. 1 are the corresponding performance metrics (available bandwidth and latency), which act as the input features to the learning algorithm.

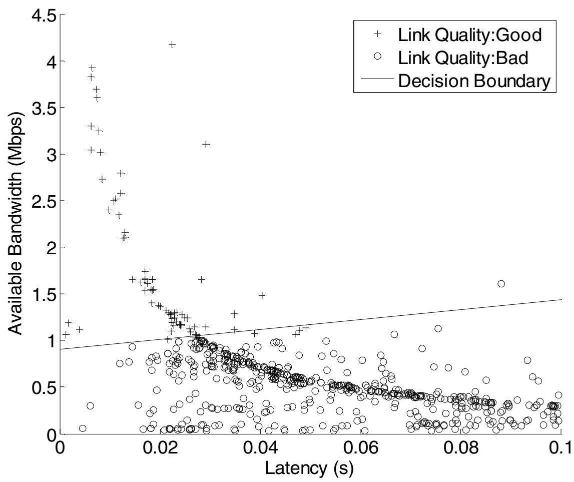

Fig. 1. Binary classification by LR

For regularized logistic regression, we have tested it with different value of regularization coefficient, λ and obtain the respective accuracy as in Table IV.

TABLE IV. BINARY CLASSIFICATION BY RLR EXPERIMENTED WITH DIFFERENT λ.

Regularization coefficient, λ	**Accuracy (%)**
	Training Set
1	97.33
10	95.33
100	92.67

The experiment where $\lambda = 1$ yields the highest accuracy as shown in Table IV. The regularization parameter is included in training set to prevent over-fitting of the algorithm. Hence, the experimental results show training accuracy for RLR is slightly lower than normal LR for binary classification.

TABLE V. BINARY CLASSIFICATION BY RLR

	Training Set	**Cross Validation Set**	**Test Set**
Accuracy (%)	97.33	96.50	98.00

We further cross validated and tested the accuracy with the trained parameter from training set where $\lambda = 1$ and the results are shown in Table V. The decision boundary generated for RLR in Fig. 2 produces a straight line from the training model.

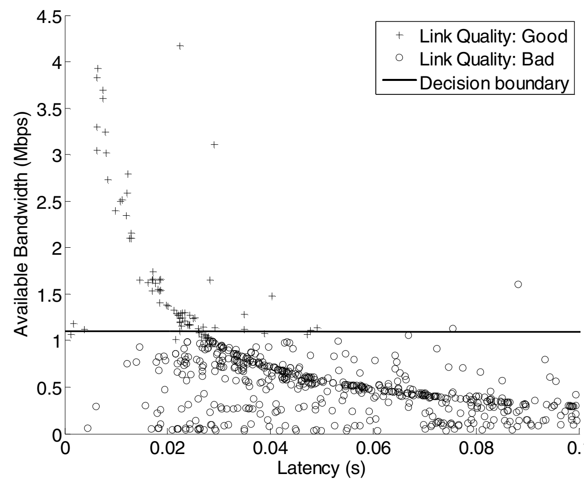

Fig. 2. Classification by RLR with λ=1.

We have implemented the multiclass regularized logistic regression through "one-against-all" method. The experimental results for multiclass classification by RLR as presented in Table VI. The illustration of multiclass classification by RLR is plotted in Fig. 3.

TABLE VI. MULTICLASS CLASSIFICATION BY ONE-AGAINST-ALL RLR EXPERIMENTED WITH DIFFERENT λ.

Regularization coefficient, λ	**Accuracy (%)**
	Training Set
1	78.33
10	78.16
100	76.50

The results in Table VI show highest accuracy is achieved when $\lambda = 1$ for training data set. The training model is cross validated and tested as shown in Table VII.

TABLE VII. MULTICLASS CLASSICIATION BY RLR

	Training Set	**Cross Validation Set**	**Test Set**
Accuracy (%)	78.33	78.00	73.50

The accuracy for multiclass RLR as shown in Table VII is lower, with around 78% of correctly predicted samples.

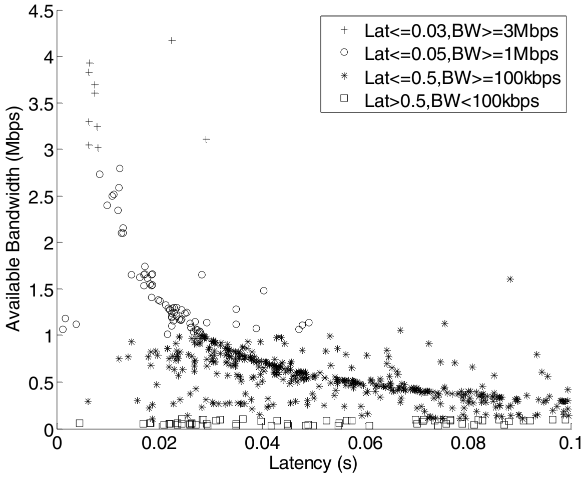

Fig. 3. Multiclass classification by one-against-all RLR with λ=1.

D. Experimental Results (C-SVC with Linear Kernel)

We have experimented with different values of the balance parameter, *C* of C-SVC as shown in Table VIII.

Table VIII. C-SVC WITH LINEAR KERNEL EXPERIEMENTED WITH DIFFERENT VALUES OF *C*

log2c	Training Set Accuracy (%)
-10	86.67
0	97.50
5	99.00
10	99.50
15	99.67
20	99.67

The experimental results show that $C=2^{20}$ is the best fit for the classification. It is known that large value of *C* will lead to over-fitting of the algorithm for smaller data size. To prevent this, the parameter *C* is set to 2^5 and applied in cross-validation set and test set.

Table VIIII. EXPERIMENTAL RESULTS FOR C-SVC WITH LINEAR KERNEL

	Training Set	Cross-validation Set	Test Set
Accuracy (%) (Binary)	99.00	99.50	98.50
Accuracy (%) (Multiclass)	98.83	98.00	99.00

From Table VIIII, C-SVC is able to achieve high accuracy in binary classification compared to multiclass classification. We have further cross-validated the respective training model obtained, which achieved 98% of accuracy.

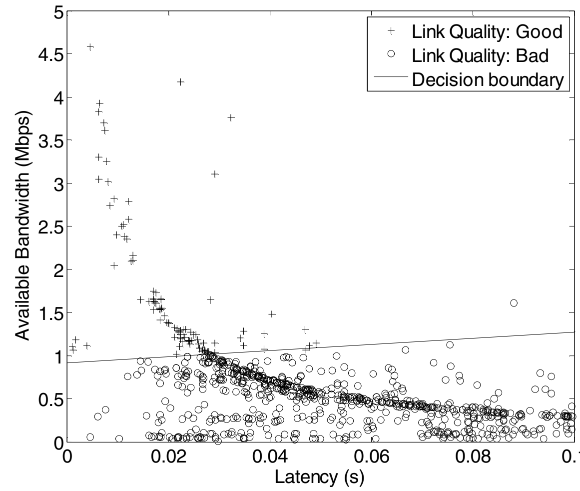

Fig. 4. Decision boundary plotted for C-SVC with linear kernel.

The decision boundary for C-SVC with linear kernel is plotted as in Fig. 4. It is shown that in Fig 4 the training data is not linearly separable, thus we applied kernel function (RBF kernel) to divide the data in a higher dimensional space in the following section.

E. Experimental Results (C-SVC with RBF kernel)

Since there are two important parameters (C, γ) in C-SVC with RBF kernel, we implemented grid search to find the best suited parameter and the result is shown in contour plot in Fig. 5 and Fig. 6.

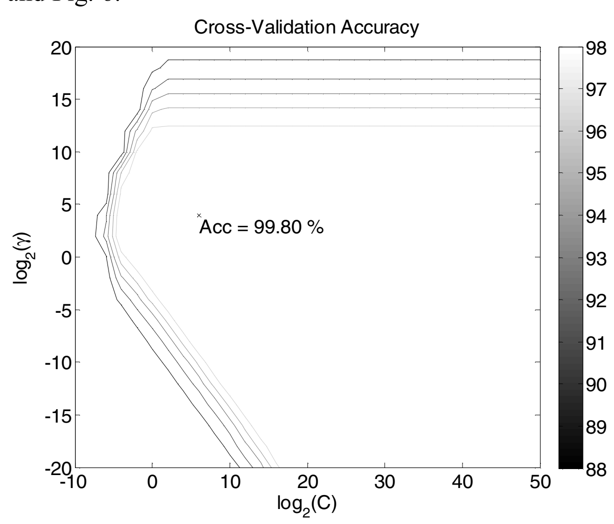

Fig. 5 Contour of cross-validation accuracy via grid-search for binary classification with SVM.

The best cross-validation accuracy is achieved at $C = 2^6$ and $\gamma = 2^4$ for binary classification while $C = 2^{12}$ and $\gamma = 2^2$ for multiclass classification. Then, the parameters (C, γ) will be used to train a model and test it again with test set.

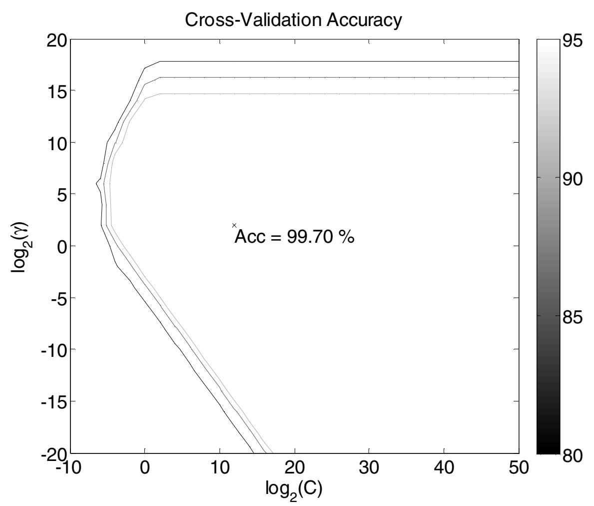

Fig. 6 Contour of cross-validation accuracy via grid-search for multiclass classification with SVM.

Table X. EXPERIMENTAL RESULTS FOR C-SVC WITH RBF KERNEL

	Training Set	Cross-validation Set	Test Set
Accuracy (%) (Binary)	99.17	99.80	99.50
Accuracy (%) (Multiclass: one-against-all)	99.83	99.70	100.00

C-SVC with RBF kernel is able to achieve high accuracy for both binary and multiclass classification, as shown in Table X. The accuracy for test set in multiclass classification gives 100% accuracy with all the sample links are correctly classified. This can be over-fit as well since the test set is small but *C* is set to be very large.

F. Discussions

We have implemented both binary and multiclass classification for logistic regression with the results shown in Section VI. The training parameter obtained from the training samples produces high accuracy for both the test set and the cross-validation set in binary classification. To further verify that the prediction algorithm does not over-fit, regularization coefficient, λ is included. We have tested it with different values of λ. As shown in Table IV, the accuracy is best achievable at $\lambda = 1$. While for multiclass classification with RLR, the accuracy is not comparable with C-SVC, with highest achievable accuracy of 78.33%.

For binary classification with C-SVC linear kernel, the regularization parameter, *C*, is experimented with different values, and shown high accuracy when $C=2^{20}$ in Table VIII. In order to prevent algorithm from over-fitting, *C* is set to 2^5. While for RBF kernel, we have performed the parameter selection through grid search, which provides high accuracy on training, cross-validation and test set as shown in Table X. It is observed that test set in multiclass classification achieved 100% accuracy. This is due to the large *C* trained from the training data over-fit the algorithm with smaller samples. C-SVC with RBF kernel gives higher accuracy compared to C-SVC with linear kernel. From the results obtained, we are able to prove that SVM outperforms logistic regression in both binary (with slightly higher accuracy) and multiclass classification (with average of 20% higher accuracy).

The training parameters obtained in the experiments are not generalizable as it depends highly on the nature of the dataset, including the network environment and number of network nodes.

G. LR vs. SVMs

The difference between LR and SVMs is that LR predicts a link's quality based on probabilistic function, while SVMs are statistical learning theory of finding a predictive function based on training data. LR is a regression analysis which maximizes the likelihood of data iteratively. SVM generates a model function and directly maximize the accuracy [5]. Kernelized SVM works better than linear SVM for non-linear separable data. LR does not require tuning parameter, except the regularization parameter in RLR to prevent over-fitting, but it is unable to achieve comparable accuracy in multiclass classification. While SVM is able to achieve high accuracy with small training dataset size and outperform LR especially in multiclass classification.

The metrics pair is constructed via our previous prediction algorithm [18], in which prediction is performed in a large scale network by measuring to a few and predicts the rest. This provide us with *O(nxn)* look-up table with *n* number of network links rather than depend solely on one single link. As network metrics varied over time, the future link condition can be predicted via the updated metrics value from the network resource prediction algorithm [18]. In real implementation, these classification approaches are integrated into the video conferencing system to suggest the supported call type to user based on the link quality.

VII. CONCLUSIONS

Knowledge of network performance metrics such as available bandwidth and latency are important for the scalability and Quality of Service in multimedia streaming. Existing prediction approaches involve only either available bandwidth or latency with real values. By taking available bandwidth and latency into consideration for link quality prediction, we are able to predict the quality of the link and suggest supported conference call type to the users. In this paper, we have implemented two major classification approaches, i.e. logistic regression and SVM to perform the classification task. Logistic regression is easier to be implemented with only one parameter, the regularization coefficient. SVM is able to give high accuracy even with small training dataset. Furthermore, it is also suitable for large classification problem since it is able to classify the data in higher dimensions with the use of kernel function. Through our experiments, SVM outperforms logistic regression in both binary and multiclass classification. In the future, we plan to experiment classification through SVM in practical use for multimedia streaming and also the online learning (real-time learning).

ACKNOWLEDGMENT

We would like to thank Boon Ping for her advice on the project and also the developers of LIBSVM for their useful toolkit and hard work.

REFERENCES

[1] B. Ratner, "Logistic Regression: The Workhorse of Response Modeling," in *Statistical and machine learning data mining: Techniques for better predictive modeling and analysis of big data*, 2012, ch. 8.

[2] M. W. Chang, W. T. Yih, C. Meek, "Partitioned logistic regression for spam filtering," in Proc. Of the 14th ACM SIGKDD Int. conf. on Knowledge discovery and data mining , pp. 97-105, 2008.

[3] M. Chae, S. Shim, H. Cho, B. Lee, "An empirical analysis of fraud detection in online auctions: credit card phantom transaction," in Proc. of the 40th Hawaii Int. Conf. on System Science, 2007.

[4] D. Timmerman et. al, "Logistic regression model to distinguish between benign and malignant adnexal mass before surgery: A multicenter study by the Int. ovarian tumor analysis group," in J. of Clinical Oncology, 2005.

[5] C. Campbell and Y. Ying, "Learning with support vector machines," in *Synthesis lectures on artificial intelligence and machine learning*, 2011, ch. *1*, pp. *1-14*.

[6] A. Este, F. Gringoli, L. Salgarelli, "Support vector machines for TCP traffic classification," in Int. J. of Comp. and Telecommun. Networking 2009.

[7] T. Simon, and D. Koller, "Support vector machine active learning with applications to text classification," in J. of Machine Learning Research, Stanford Univ, USA, 45-66, 2001.

[8] T. S. Furey, N. Cristianini, N. Duffy, et al., "Support vector machine classification and validation of cancer tissue samples using microarray expression data," in BIOINFORMATICS, USA, vol. 16, pp. 906-914, May 2000.

[9] F. Dabek, R. Cox, R. Kaashoek, and R. Morris, "Vivaldi: A decentralized network coordinate system," in SIGCOMM'04, USA, August 2004.

[10] T. S. Eugene Ng and H. Zhang, "Predicting internet network distance with coordinates-based approaches," in Proc. of IEEE Infocom'02, New York, June 2002.

[11] Y. Mao, L. Saul, and J. M. Smith, "IDES: An internet distance service for large networks," in IEEE J. on Selected Areas in Comm., pp. 2273-2284, December 2006.

[12] O. Beaumont, L. Eyraud-Dubois, and Y. J. Won, "Using the last-mile model as a distributed scheme for available bandwidth prediction," in EuroPar 2011, France, April 2011.

[13] V. Ramasubramanian, D. Malkhi, F. Kuhn, M. Balakrishnan, et al., "On the treeness of internet latency and bandwidth," in SIGMETRICS/Performance'09, USA, June 2009.

[14] S. Schober, S. Brenner, R. Kapitza, et al., "Bandwidth prediction in the face of asymmetry," in Proc. 13th Int. IFIP Conf. on Distrib. Appl. And Interop. Sys., January 2013.

[15] Y. J. Liao, P. Geurts and G. Leduc, "Network distance prediction based on decentralized matrix factorization," in Proc. of IFIP Networking 2010, India , May 2010.

[16] Y. Liao, W. Du, P. Geurts, and G. Leduc, "DMFSGD: A decentralized matrix factorization algorithm for network distance prediction," in IEEE/ACM Trans. on Netw., 2012.

[17] W. Du, Y. Liao, P. Geurts and G. Leduc, "Ordinal rating of network performance and inference by matrix completion," Available: http://arxiv.org/abs/1211.00447.

[18] S. J. Lim, B. P. Lim, S. Lau, et al., "Implementing stochastic gradient descent based on historical network distance for available bandwidth prediction," in 2nd Int. Conf. on e-Technologies and Network for Development, 2013.

[19] P. Komarek, "Logistic regression for data mining and high-dimensional classification," M.S. thesis, Dept. of Math Sciences, Camegie Mellon Univ.

[20] V. N. Vapnik, "An overview of statistical learning theory," in IEEE Transaction on Neural Networking, September 1999.

[21] B. E. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in Proc. of the 5th Annual Workshop on Computational Learning Theory, ACM Press, pp. 144-152, 1992.

[22] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," in ACM Trans. on Intelligent Systems and Technology*,* 2001.

[23] L.Eyraud-Dubois and P. Uznanski, "Bedibe: Datasets and Software Tools for Distributed Bandwidth Prediction,", in *AlgoTel*, 2012.

[24] Skype homepage. [Online]. Available: https://support.skype.com/en/faq/FA1417/how-much-bandwidth-does-skype-need

[25] TrueConf homepage. [Online]. Available: http://trueconf.com/support/communication-channels.html

[26] G. Karlson, "Asynchronous transfer of video," in *IEEE Communications Magazine*, August 1996, pp. 118-126.