Performance Evaluation of a Decision-Theoretic Approach for Quality of Experience Measurement in Mobile and Pervasive Computing Scenarios

Karan Mitra[†‡]
†Faculty of Information Technology
Monash University
900 Dandenong Road, Caulfield East
Victoria, Australia, 3145
Email: karan.mitra@monash.edu

Christer Åhlund[‡]
‡ Luleå University of Technology
SE-97187, Luleå, Sweden
Email: christer.ahlund@ltu.se

Arkady Zaslavsky[‡§]
§ICT Centre, CSIRO, Acton
ACT, Australia, 2601
Email: arkady.zaslavsky@csiro.au

Abstract—**Measuring and predicting users quality of experience (QoE) in dynamic network conditions is a challenging task. This paper presents results related to a decision-theoretic methodology incorporating Bayesian networks (BNs) and utility theory for quality of experience (QoE) measurement and prediction in mobile computing scenarios. In particular, we show how both generative and discriminative BNs can be used to measure and predict users QoE accurately for voice applications under several wireless network conditions such as wireless signal fading, vertical handoffs, wireless network congestion and normal hotspot traffic. Through extensive simulation studies and results analysis, we show that our proposed methodology can achieve an average accuracy of 98.70% using three different types of Bayesian network.**

I. INTRODUCTION

Quality of experience (QoE) refers to users likes/dislikes towards a particular service/object/tool used by them. It is dependent on users experience and expectations and their cognitive, behavioural and psychological states which are not observed directly. However, these states do dictate their QoE [2], [19], [17], [12]. In mobile and pervasive computing environments, users use their mobile devices such as smartphones on-the-move where situations related to them and the underlying network conditions can be highly dynamic and uncertain. For example, if a user is in a bus, his/her QoE might be lower as compared to when they are at home. This can be caused by the noisy environment or perhaps a number of people nearby.

There is a need to develop QoE measurement and prediction approaches that can consider several parameters (both subjective and objective) under different situations related to users and network connectivity alike [2], [19], [11]. Once developed, these approaches can help stakeholders such as network operators to maximize their revenues by understanding customers QoE of services being offered to them and thereby minimizing network churn. These approaches can also be used to provide users with personalized services on their mobile devices.

Recently, Mitra *et al.* [11], proposed a context-aware decision-theoretic approach for QoE measurement and predic-tion in mobile and pervasive computing environments. However, they do not consider the problem of QoE measurement and prediction under multiple diverse network conditions such as wireless signal fading, handoffs and network congestion. Thus, this paper presents the following contributions: Firstly, we carefully study the QoE related to VoIP application in different scenarios and under several network conditions such as, wireless signal fading, vertical handoffs, wireless network congestion and normal background network traffic. Finally, we show how simple BNs are able to correctly measure and predict users QoE in these scenarios.

This paper is organized as follows: Section II presents the related work. Section III presents our approach for QoE measurement and prediction. Section IV presents the results analysis. Finally, section V presents the conclusion and future work.

II. RELATED WORK

Recently, Moller *et al.* [14] and Tsompanidis *et al.* [18] consider the problem of QoE measurement for VoIP applications in heterogeneous access networks by considering network impairments such as handoffs and network congestion. However, they consider models such as the ITU-T E-Model [4] for QoE measurement. These models can be inefficient when measuring QoE in mobile computing scenarios where network QoS and users situation can be highly dynamic. Further, these models cannot deal with uncertainty, biased and missing information which is typical for human judgements [12].

Sung [17], developed algorithms and architectures for QoE reasoning and adaptation. However, their approach was limited to mobile video applications. Wu *et al.* [19] defines and classify the QoE related parameters and try to find correlation between QoE and QoS related classes. Their approach can be impractical when there are several QoE and QoS parameters as finding correlations between each and every parameter is a complex task.

Moller *et al.* [13], develops a taxonomy for QoE based on which stakeholders such as application designers can select

QoE and QoS parameters to develop multimedia applications. However, they do not present methods for modelling QoS and QoE relationships. Brooks and Hestnes [2] consider the problem of combining subjective and objective methods to determine overall QoE. Mitra *et al.* [11] considers aforementioned problems of defining interdependence of multiple QoE and QoS parameters to model, measure and predict overall QoE based on a decision-theoretic approach. However, they do not measure users' QoE under several network impairments.

III. A DECISION-THEORETIC APPROACH FOR QUALITY OF EXPERIENCE MEASUREMENT AND PREDICTION

Relationships between several parameters such as delay, jitter, packet loss and users QoE has been studied widely in the literature [14], [18]. Even then, these relationships may not correctly model users QoE due to dynamic nature of underlying network conditions and the user environment where uncertainty plays a dominant role. Further, users QoE depend on context parameters such as people nearby, location of the user, their moods, and their stress level [3], [10], [13], [17]. This problem can become even more difficult when more parameters for QoE measurement are considered. The relationships between these parameters can be complex and defining a suitable mathematical model can be quite challenging.

It is also important to note that user ratings are subjective and are measured on ordinal scale such as the 5 or 7 point Likert-like scales. Thus, performing any mathematical operation using these scales directly will be incorrect as the alternatives on these scales are rank ordered and the distance between each alternative cannot be determined [7], [11]. Thus, techniques like linear regression cannot be applied directly. Using the approach presented by Mitra *et al.* [11], researchers can solve such problems. The authors show that by using BNs and utility theory, QoE can be measured efficiently on a single scale. Further, they also provide approaches for QoE learning and prediction. We now use BN for QoE measurement and prediction in multiple diverse network conditions which are prevalent in mobile computing environments.

A BN is a directed acyclic graph (DAG) where random variables form the nodes of a network. The directed links between the nodes form the causal relationships. The direction of a link from node X to node Y means that X is the parent of Y. Any entry in the network can be calculated using the joint distribution denoted as [16]:

$$P(x_i, ..., x_n) = \prod_{i=1}^{n} P(x_i | Parents(X_i)) \qquad (1)$$

where, nodes $Parents(X_i)$, are the parents of node x_i. At the lowest level, context information/attributes ($a_n^t \in A$) such as bandwidth ($a_{bandwidth}^t$) and location ($a_{location}^t$) are collected to infer context states such as technology acceptance (S_{TA}^t) and user satisfaction (S_{US}^t) at the next level. These states are *"hidden"* as we don't know users intention and should be inferred directly from the observed context attributes.

A BN for QoE measurement can be created in many ways. We can create a BN for QoE measurement by creating 'arcs' between all these parameters and learn the network parameters from the collected data. Using e.q.1, any query in the BN can be answered in a straightforward manner. By using inference algorithms such as variable elimination and junction tree [16], we can determine hypotheses about users QoE in an efficient manner.

Fig. 1(a) and 1(b) shows two BNs for voice quality assessment based on noisy-OR model and Naive Bayes model [8], [16]. One of the simplest BN is the Naive Bayes network (NBN) as shown in fig. 1(b) where all the observations or context attributes are conditionally independent given an outcome such as QoE is "excellent". Thus, we have a generative model in the form of:

$$P(R_{QoE}^t | a_n^t) = P(R_{QoE}^t) \prod_n P(a_n^t | R_{QoE}^t) \qquad (2)$$

In some cases, the independence assumptions between the parameters to determine a hypotheses may be too strict. In this model, the observations are assumed to the discrete i.e., a_{delay}^t, a_{jitter}^t and $a_{packetlosses}^t$ are discretized into finite states. Another form of NBN is the Gaussian Naive Bayes (GNB) which assumes the observations or context attributes are represented by Gaussian distribution. This form of model can handle continuous observations directly without the need for discretization.

Another simple model as shown in fig. 1(a) is the noisy-OR model or the discriminative model. This model is used to predict the state directly from the given observations or $P(R_{QoE}^t | a_n^t)$. For a hybrid BNs (HBN) with continuous and discrete nodes, one has to use the softmax functions to perform inference. This model can be written as:

$$P(R = r_i | a_1^t, ..., a_n^t) = \frac{exp(b^i + \Sigma_{l=1}^n w_l^i a_l^t)}{\Sigma_{j=1}^m exp(b^j + \Sigma_{l=1}^n w_l^j a_l^t)} \qquad (3)$$

E.q. 3 defines a softmax model, which is similar to performing multinomial logistic regression when continuous nodes (a_n^t) are the parents of the discrete nodes (R). An important property here is that in a softmax model, stakeholders can add domain knowledge which is not possible in simple logistic regression. In this model, parent nodes are defined usually as Gaussian distribution for continuous attributes such as $a_{bandwidth}^t$ and conditional probability tables (CPT) are defined for discrete nodes such as $a_{location}^t$. The CPDs of the children nodes such as S_{US}^t is defined as the softmax nodes as S_{US}^t is measured on the ordinal scale. There are various other ways to construct a BN which is out of the scope of this paper. Indeed, we will show that how these simple models can measure QoE very accurately.

For overall QoE measurement based on multiple QoE states such as S_{TA}^t and S_{US}^t, firstly, a Bayesian network (BN_1) is created and learned based on the $initial_{dataset}$ which contains observed context attributes and user provided ratings. Then the QoE for all states such as S_{US}^t and S_{TA}^t is inferred based on e.q.1, 2 or 3. These states are then mapped onto single
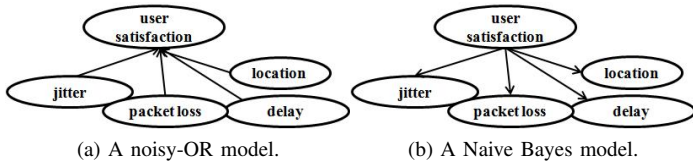
(a) A noisy-OR model. (b) A Naive Bayes model.

Fig. 1: Bayesian networks for QoE measurement.

(a) Fading. (b) Vertical handoff.

(c) Network congestion. (d) Normal hotspot traffic.

Fig. 2: Multiple scenarios which impact QoE related to multimedia applications such as VoIP.

utility values. These utility values are then added together as a weighted linear combination to determine a single global utility which is the final QoE value.

For correct QoE estimation, Mitra *et al.* [11] consider Bipolar interval scale [7] such that the outcomes of states S_{US}^t such as "excellent" and "poor" are equidistant from each other. The highest achievable utility value, '1', is mapped to the best possible QoE outcome such as "excellent" and lowest utility value, '0', is mapped to worst possible outcome say, "poor". Similarly, they then assign utilities to other outcomes such as "very good" , "good" and "fair", as '0.75', '0.50' and '0.25', respectively (assuming 5 point Likert-like scale. Indeed other scales can be used). This way, subjective measurement outcomes in the form of hypotheses and utilities can be mapped on to the objective scale. Consider an example, where 9 out of 10 users' give QoE ratings for a VoIP call as "excellent" i.e., $p(S_{US}^t = \text{``excellent''}) = 0.90$ and 1 user gives the outcome "very good" i.e., $p(S_{US}^t = \text{``verygood''}) = 0.10$ for context state, S_{US}^t. The expected utility of the context state will be $EU(S_{US}^t) = [(1*0.90)+(0.75*0.10)] = 0.9750$. This, utility value is then mapped onto the Bipolar interval scale to determine the QoE of S_{US}^t. As $EU(S_{US}^t)$ closer to 1, it means overall QoE is "excellent".

Similarly, expected utility for other context states such as technology acceptance $EU(S_{TA}^t)$ can be determined. These states are then added as a weighted linear combination to determine the single global QoE utility value $GU(QoE)$ for the overall QoE. $GU(QoE)$ is again mapped on to the Bipolar scale to determine the overall QoE. Once the overall QoE is determined for all the cases in the $initial_{dataset}$, these are then appended as a new column in $initial_{dataset}$ and is renamed as the $append_{dataset}$. A second BN (BN_2) is then created and is re-learned based on $append_{dataset}$ to make correct predictions on the new data. In order to accurately measure and predict overall QoE based on BN_1 and BN_2, both BNs should correctly learn and infer QoE for all the states. We will now determine whether the BNs can correctly learn and make correct inference related to QoE in mobile computing scenarios. In this paper, we assume to have a single QoE class i.e, user satisfaction (S_{US}^t).

IV. RESULTS ANALYSIS

For results analysis, we considered several scenarios such as wireless signal fading, wireless network congestion, vertical handoffs and normal hotspot traffic in IEEE 802.11 WLAN network environments (as shown in fig.2). We considered VoIP application that can use both ITU-T G.711 and ITU-T G.729
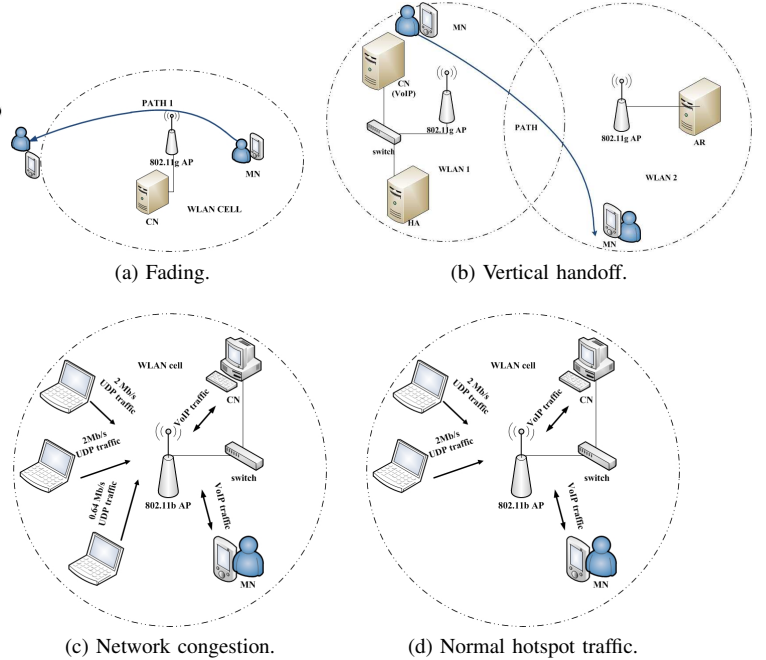
voice codecs. To determine user satisfaction (S_{US}^t) in terms of the MOS mapped on the scale of 1 to 5, we considered the ITU-T E-Model [4]. Where '1' means "poor", '2' means "fair", '3' means "good", '4' means "very good" and '5' means "excellent", respectively. Indeed, our model competes directly against the ITU-T E-Model but in order to test the measurement and prediction capability of our proposed model under diverse network conditions, we find the output of ITU-T E-Model to be sufficient for our purposes as it consider non-linearity in delay and packet loss impairments to output the MOS. Further, using OPNET simulator [1], we can generate different datasets for several scenarios based on realistic simulation setup and assumptions. Using our studies, we gather that our approach is both space and time efficient and can be used in scenarios where correct QoE measurements are required on-the-fly. The applications do not have to perform expensive calculations to obtain MOS using the ITU-T E-Model yet resulting in significant QoE prediction accuracy.

A. Performance Evaluation of VoIP applications in Mobile and Pervasive Computing Scenarios

1) Wireless signal fading in IEEE 802.11g WLAN: In case of wireless signal fading in IEEE 802.11g WLAN, a mobile node (MN) moved out of the coverage area of an access point (AP) while having a voice conversation with the correspondent node (CN). In this scenario, we only wanted to determine the effects of mobility on the voice conversation. Thus, we do not consider any background network traffic. The length of a call was set to 30 seconds after which the user moved out of the coverage area of the AP and the call was dropped. Fig. 2(a)

TABLE I: Cumulative results related to all the scenarios for ITU-T G.711 voice codec.

Scenario	Avg. delay (ms)	Avg. PLR (%)	Avg jitter (ms)	Avg. E-Model MOS
Network congestion	414	0.03	0	2.06
Signal fading	93	0	0.06	3.59
Vertical handoff	108	10.05	0	2.75
Normal hotspot traffic	71	0	0	3.62

TABLE II: Cumulative results related to all the scenarios for ITU-T G.729 voice codec.

Scenario	Avg. delay (ms)	Avg. PLR (%)	Avg. jitter (ms)	Avg E-Model MOS
Network congestion	170	0	0	2.41
Signal fading	92	0	0.02	3.00
Vertical handoff	108	7.61	0	2.50
Normal hotspot traffic	70	0	0	3.01

shows the targeted scenario.

Fig. 3 shows the effects of mobility on the MOS for the ITU-T G.729 codec. As soon as the user reached a certain signal to noise ratio (SNR) threshold (14 dB), the MOS suddenly dropped from around 3 to 1. This is also observed for ITU-T G.711 codec. The cause of the sudden drop in the MOS can be attributed to the sudden rise in network delay due to extremely low SNR. As can be seen in fig. 3, the end-to-end delay increased suddenly from 0.060 milliseconds (ms) approx. to 1 second (sec) approx. for ITU-T G.729 codec case.
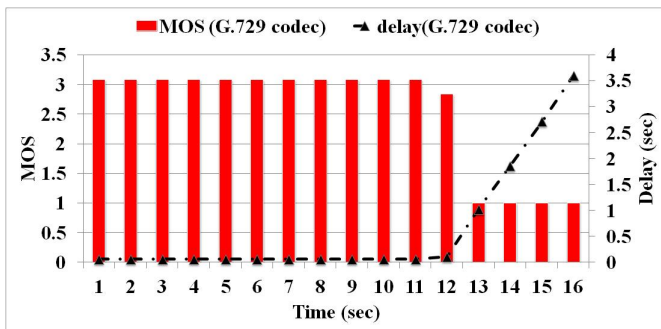

Fig. 4: MOS related to ITU-T G.711 codec in case of vertical handoffs.

Fig. 3: Observed MOS related to ITU-T G.711 and ITU-T G.729 codecs in case of wireless signal fading in 802.11g WLAN.

2) Vertical handoffs in vehicular networks using Mobile IPv6: In this case, we considered vertical handoffs between two WLAN networks at low vehicular speeds (10 km/hr). Fig. 2(b) shows the targeted scenario. We considered Mobile IPv6 protocol [9] for mobility management. A vehicle carrying a MN moved from one WLAN network to another. In this case, average delay caused due to handoff was quite low at 108 ms with average packet loss ratio of 10.05% and 7.61% for ITU-T G.711 and ITU-T G.729 codecs, respectively. These statistics are fairly typical in case of hard handoffs where connection from one network is lost so that a MN can attach itself to a new network. Fig. 4 shows the effects of hard handoff on the MOS for ITU-T G.711 codec. In both cases, we assume that handoff occurs at the middle of a voice call. In reality, location of handoff within a call can have different impact in users QoE [14].
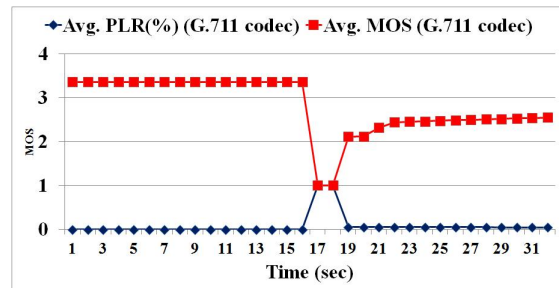
3) IEEE 802.11b WLAN network congestion : To determine the effects of network congestion on voice quality, we used the IEEE 802.11b WLAN connection. To saturate the AP, we set up three wireless nodes to generate background UDP traffic as shown in fig 2(c). The maximum achievable bit rate was approximately 5Mb/s after which the AP dropped all packets due to buffer overflow. All three nodes generated 175 packets per second towards the CN. Once the network reached its steady state, the MN initiated a call to the CN and calculated the required statistics. In this case, we set the call duration to 16 minutes and calculated the average statistics. From simulations, we gather that in case of heavy UDP traffic, network congestion occurs mainly due to increase end-to-end delay which in turn causes MOS to vary. Fig. 5 shows the fluctuation of MOS w.r.t average end-to-end delay in case of ITU-T G.729 codec. We observed that overall packet loss rate and jitter were close to 0 for both codecs. The average MOS for ITU-T G.711 codec was 2.06. For ITU-T G.729 codec, it was 2.41.

4) Normal hotspot network traffic: In this case, the MN enjoyed near perfect voice call quality with extremely low delay and jitter. Fig. 2(d) shows the targeted scenario. Two MNs generated approximately 4 Mb/s of background UDP network traffic towards the CN without saturating it. After sometime, a voice call was established between the CN and the MN. As the amount of traffic generated was not enough to cause congestion, the VoIP call enjoyed very low delays and close to zero jitter and packet losses. Hence, the average MOS

calculated was greater than 3 for both codecs. Fig. 6 shows the average MOS for the voice call using the ITU-T G.729 codec.
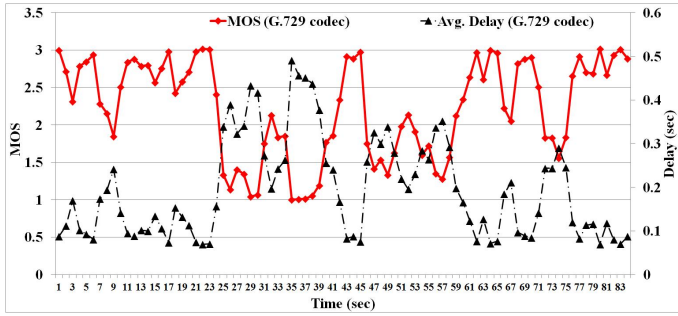

Fig. 5: Performance of ITU-T G.729 codec in presence of network congestion.
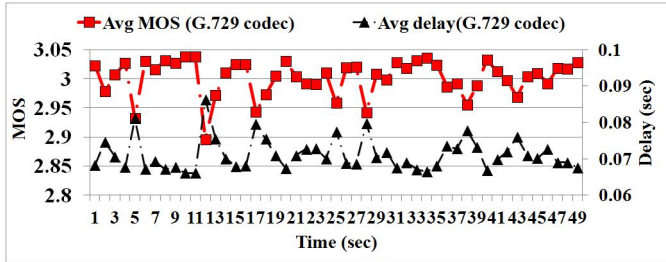

Fig. 6: Performance of ITU-T G.729 codec in presence of normal hotspot network traffic.

Table I and II shows the results related to all the scenarios for both codecs.

B. QoE Prediction and Measurement for VoIP Applications in Mobile Computing Scenarios

We developed two prototypes using the Bayes Net Toolbox for MATLAB (BNT) [15] and Weka data mining software [5]. We used BNT for the noisy-OR type hybrid Bayesian network (HBN) and we used Weka data mining software to create Gaussian Naive Bayes (GNB) and Naive Bayes (NBN) networks. Based on OPNET-based simulation studies [1], we obtained two datasets related to both ITU-T G.711 (dataset 1) and ITU-T G.729 (dataset 2) codecs. For each scenario, we collected results based on 125 simulation runs with different random seeds. We then train our BNs to measure and predict the QoE. For training the model parameters, we use the expectation maximization (EM) algorithm [16], [6]. In both datasets, we had 500 training cases (125 runs*4 scenarios) each representing results related to multiple diverse network conditions.

Tables III and IV show statistics related to all parameters and scenarios in both datasets. We can see that both datasets contains wide variations in terms of parameter values. In particular, for average end-to-end delay (a_{delay}^t), average packet losses ($a_{packetloss}^t$) and average user satisfaction (S_{US}^t). As the E-Model outputs continuous values for MOS, we discretized

these values such that the values greater than the middle point in the interval is mapped to the maximum alternative. The value smaller than the middle point in that interval is mapped to the lowest alternative. For example, a value of 2.59 is mapped to 3 and 2.25 is mapped to 2. The reason for doing this is that in reality, users give ratings on the ordinal scale and select only few alternatives on the scale. We also tried several other possible ways of discretization and achieved similar results.

TABLE III: Statistics related to all parameters in dataset 1.

Parameters	Min. value	Max. value	Mean	Standard deviation
a_{delay}^t	68 ms	671 ms	172 ms	146 ms
a_{jitter}^t	0 ms	3 ms	1 ms	1 ms
$a_{packetloss}^t$	2.5 %	25.4%	2.5%	4.6%
S_{US}^t	1	4	-	-

TABLE IV: Statistics related to all parameters in dataset 2.

Parameters	Min. value	Max value	Mean	Standard deviation
a_{delay}^t	67 ms	374 ms	109 ms	41 ms
a_{jitter}^t	-1 ms	3 ms	1 ms	1 ms
$a_{packetloss}^t$	0 %	20.5%	1.9%	3.7%
S_{US}^t	1	3	-	-

Based on the collected datasets, we train the BNs using EM algorithm [16], [6]. As the data is fully available, EM reduces to maximum likelihood estimation (MLE) [16], [6] as in the supervised learning case. As we are dealing with learning the BN parameters, the best approach for their evaluation is to perform cross validation where the idea is to estimate models prediction accuracy by keeping aside some fraction of the training data for learning. The remaining data is used as the test data for which hypothesis of the states are inferred [16]. In this paper, we have used 10-fold cross validation for checking models prediction accuracy. Table V and VI shows the accuracy of predictions for both datasets.

TABLE V: Prediction accuracy of Bayesian networks for voice quality measurement for dataset 1.

Bayesian network type	prediction accuracy (%)
Naive Bayes Network (NBN)	99.40
Gaussian Naive Bayes Network (GNB)	99.40
Hybrid Bayesian Network (HBN)	99.40

TABLE VI: Prediction accuracy of Bayesian networks for voice quality measurement for dataset 2.

Bayesian network type	prediction accuracy (%)
Naive Bayes Network (NBN)	97
Gaussian Naive Bayes Network (GNB)	97.40
Hybrid Bayesian Network (HBN)	99.60

As can be observed, all three type of BNs give excellent prediction accuracy based on 10-fold cross validation. Even these simple BNs (fig. 1(a) and fig.1(b)) are able to correctly

predict users QoE with overall average accuracy of 98.70%. In dataset 1, both generative (NBN and GNB) and discriminative (HBN) performs similarly with all three achieving an accuracy of 99.40%. In case of dataset 2, HBN performs marginally better than the GNB and NBN with an accuracy of 99.60%. NBN and GNB achieved an accuracy of 97% and 97.40%, respectively. Most importantly, in all cases, the prediction accuracy is greater than 95% which is an excellent result.

C. Discussion and lessons learnt

From our results, we can conclude that BNs can be used for accurate QoE measurement and prediction. From the implementation perspective, both discriminative and generative BNs work very well on our data sets. It might happen however that, generative or discriminative BNs might outperform each other in different cases. In such a scenario, a BN should be chosen which provides better prediction accuracy. Applications can be developed easily that can take as input the available datasets and automatically outputs the correct QoE measurements by selecting optimal BNs automatically.

Its worth noting that more objective and subjective parameters can be added and deleted as the per the stakeholders requirement and the BNs can easily accommodate such changes while maintaining low complexity and higher accuracy. Further, domain knowledge can be added in BNs which can be beneficial for the stakeholders in case they want to understand the effects of some parameters without collecting additional data. These BNs can also be applied directly for real-time applications and protocols for dynamic network selection and codec change. Inference in these models usually takes less than 50 milliseconds for NBN and GNB and around 120 milliseconds for HBN for our datasets which also include learning the model parameters.

V. Conclusion and Future Work

This paper presents performance evaluation of a decision-theoretic approach for efficient quality of experience measurement and prediction. Through extensive simulations, we study the impairments caused by diverse network conditions such as wireless signal fading, vertical handoffs and network congestion on VoIP applications. Using prototype implementations, we further show that Bayesian networks, both generative and discriminative are efficient methods for QoE measurement and prediction. Using the results presented in this paper, stakeholders such as network operators can apply our approach to correctly measure QoE related to multimedia applications efficiently. We are currently applying the insights gained in this paper to develop efficient mobility management protocols for QoE guaranties in mobile and pervasive computing environments.

References

[1] Opnet network simulator, http://www.opnet.com.

[2] P. Brooks and B. Hestnes. User measures of quality of experience: why being objective and quantitative is important. *Network, IEEE*, 24(2):8–13, march-april 2010.

[3] K. De Moor, I. Ketyko, W. Joseph, T. Deryckere, L. De Marez, L. Martens, and G. Verleye. Proposed framework for evaluating quality of experience in a mobile, testbed-oriented living lab setting. *Mob. Netw. Appl.*, 15:378–391, June 2010.

[4] ITU-T Recommendation G.107. Itu-t recommendation g.107, methods for subjective determination of transmission quality. 2008.

[5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *SIGKDD Explor. Newsl.*, 11:10–18, November 2009.

[6] D. Heckerman. A tutorial on learning with bayesian networks. In *Proceedings of the NATO Advanced Study Institute on Learning in graphical models*, pages 301–354, Norwell, MA, USA, 1998. Kluwer Academic Publishers.

[7] C. L. Hwang and K. P. Yoon. *Multiple attribute decision-making: Methods and applications*. Springer-Verlag, 1981.

[8] F. V. Jensen. *Bayesian Networks and Decision Graphs*. Statistics for Engineering and Information Science. Springer, 2001.

[9] DB. Johnson, C. E. Perkins, and J. Arkko. Mobility support in ipv6 (rfc 3375). http://www.ietf.org/rfc/rfc3375.txt, June 2004.

[10] S. Jumisko-Pyykkö and M. M. Hannuksela. Does context matter in quality evaluation of mobile television? In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, MobileHCI '08, pages 63–72, New York, NY, USA, 2008. ACM.

[11] K. Mitra, C. Åhlund, and A. Zaslavsky. A decision-theoretic approach for quality-of-experience measurement and prediction. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pages 1–4, july 2011.

[12] K. Mitra, A. Zaslavsky, and C. Åhlund. A probabilistic context-aware approach for quality of experience measurement in pervasive systems. In *Proceedings of the 2011 ACM Symposium on Applied Computing (SAC)*, 2011.

[13] S. Moller, K.-P. Engelbrecht, C. Kuhnel, I. Wechsung, and B. Weiss. A taxonomy of quality of service and quality of experience of multimodal human-machine interaction. In *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pages 7–12, 29-31 2009.

[14] S. Moller, M. Waltermann, B. Lewcio, N. Kirschnick, and P. Vidales. Speech quality while roaming in next generation networks. In *Communications, 2009. ICC '09. IEEE International Conference on*, pages 1–5, 2009.

[15] K. Murphy. The bayes net toolbox for matlab. In *Comput Sci Stat*, volume 33, 2001.

[16] S. Russel and P. Norvig. *Artificial Intelligence: A modern apporach*. 2 edition, 2006.

[17] J. Sun. *Football on mobile phones: algorithms, architectures and quality of experience in streaming video*. PhD Thesis, Umeå University, 2006.

[18] I. Tsompanidis, G. Fortetsanakis, and M. Papadopouli. Measuring the perceived quality of voip under various wireless network conditions [technical report]. 2010.

[19] W Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, and Z. Yang. Quality of experience in distributed interactive multimedia environments: toward a theoretical framework. In *MM '09: Proceedings of the seventeen ACM international conference on Multimedia*, pages 481–490, New York, NY, USA, 2009. ACM.