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

Karan Mitra[†‡], *Christer Åhlund*[‡], *Arkady Zaslavsky*[†‡]

[†]Faculty of Information Technology, Monash University, Melbourne, Australia
[‡] Luleå University of Technology, SE-97187, Luleå, Sweden

ABSTRACT

This paper presents a pioneering context-aware approach for quality of experience (QoE) measurement and prediction. The proposed approach incorporates an intuitive context-aware framework and decision theory. It is capable of incorporating several QoE related classes and context information to correctly measure and predict the overall QoE on a single scale. Our approach can be used in measuring and predicting QoE in both lab and living-lab settings based on user, device and network related context parameters. The predicted QoE can be beneficial for network operators to minimize network churn and can help application developers to build smart user-centric applications. We perform extensive experimentation and the results validate our approach.

***Index Terms*—** Bayesian network, context-awareness, decision theory, quality of experience (QoE), quality of service (QoS)

1. INTRODUCTION

Quality of experience (QoE) of a user regarding a particular service or object not only depends on underlying quality of service (QoS) parameters but also on his/her personal and environmental context [1, 2]. Recently, approaches for QoE measurement and prediction have been proposed. For example, Wu *et al.* [1], define and classify QoE related parameters and try to find correlation between QoE and QoS related classes. This approach can be impractical when there are several QoE and QoS parameters as finding correlations between each and every parameter is a complex task. Brooks and Hestnes [2], stress the need to consider a combination of subjective and objective methods to determine QoE. Based on the state-of-the art, we gather that a unified approach for QoE measurement and prediction is missing. Also, techniques that can simultaneously handle user-centric context and subjective and objective assessment metrics are required.

In order to build a context-aware approach for QoE measurement and prediction, there are several challenges that need to be considered. Firstly, there can be complex and inter-dependent relationships between several QoE classes and parameters from user, network, device and the environment surrounding the user. These relationships can be difficult to determine analytically or mathematically. Secondly, context information can be ambiguous, biased and missing such as, user provided quality ratings and information collected from sensors and network probes. Finally, there is a need to measure QoE on a single scale [2].

To tackle these challenges, in this paper, we present a novel context-aware approach for measuring and predicting QoE. In particular, 1.) we present a way to handle complex inter-dependencies between context parameters and QoE classes; 2.) we present a method for dealing with uncertainty in measuring QoE based on human judgments; and 3.) we propose a method to map several QoE classes onto a single scale. For challenges 1 and 2, we consider context spaces theory (CST) [3] for context modelling. Bayesian networks (BN) [4] and parameter learning are considered for QoE measurement and prediction. For challenge 3, we use BN, utility theory and a bi-polar scale to map and measure multiple QoE classes into a single QoE value.

2. CONTEXT-AWARE QUALITY OF EXPERIENCE MEASUREMENT AND PREDICTION

In this section, we present our context-aware, decision-theoretic based approach to measure and predict QoE. Figure 1 shows our BN. At the lowest level, we collect context information (a_n^t) such as bandwidth ($a_{bandwidth}^t$) and location ($a_{location}^t$). At the intermediate level, context states ($S_{QoE\_state}^t$) such as technology acceptance (S_{TA}^t) and user satisfaction (S_{SA}^t) are inferred probabilistically. The top-most state is the situation state or the goal state (R_{QoE}^t) which determines the overall QoE of the user and is dependent on the context states. Figure 3 shows our algorithm to determine and predict QoE. QoE is determined in the first pass and is predicted in the second.

Firstly, we need to define the range of each context attribute and context state related to QoS, QoE, the user and his/her device environment. Based on decision theory, we also need to define the decision alternatives ($a_n \in A$) and hypotheses ($h_n \in H$). $P(\bullet)$ represents the belief of the agent in a hypothesis. Utility, $U(\bullet)$ encodes the preference on a
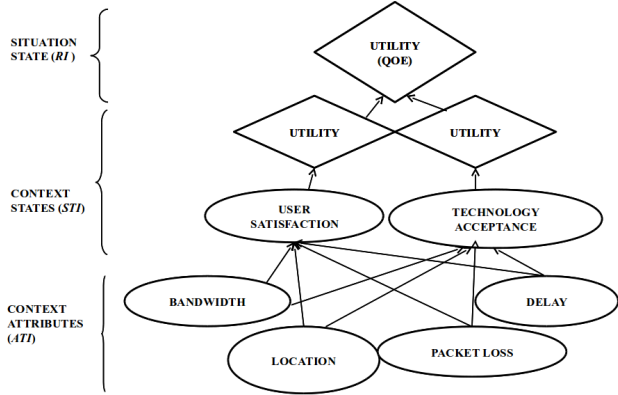
Fig. 1. Bayesian network and utility nodes to measure QoE.

Fig. 2. The bi-polar interval scale.

numerical scale. In this paper, for $h_n = \{$"excellent", "very good", "good", "fair" and "poor"$\}$, we assign $U(h_n) = \{1, 0.75, 0.50, 0.25, 0\}$. These utility values can help the stakeholders (such as network operators and application developers) to decide the importance of each decision alternative, a on the scale of 1 to 5. Formally, for each hypothesis, h_n, we have a $U(h_n, Pr)$ describing the utility if H is in state h_n. The expected utility (EU) for an individual QoE class or context state ($S_{QoE\_state}^t$) is:

$$EU(S_{QoE\_state}^t) = \sum_{h_n \in H} U(h_n, pr) P(h_n) \qquad (1)$$

Once the utility for each context state (QoE class) is calculated using e.q.1, we map the utility on a scale to select a decision alternative, a. Since we are dealing with human judgments and ratings, we need to consider a proper scale of measurement. Typical scales that are used are the ordinal or the interval scales [5, 2]. Dealing with the interval scale in case of human judgments can be difficult, therefore mostly ordinal scale is used. For example, the mean opinion score (MOS) is calculated based on the ordinal scale. This scale ranks the choices or alternatives and thus, meaningful results such as average, standard deviation and ratio cannot be derived [2]. This is due to the fact that distance between the alternatives are not fixed. To solve this problem, we propose to use the bi-polar scale [5] as shown in figure 2.

By using a bi-polar scale, we can map qualitative alternatives onto an interval scale [5]. For it's proper calibration, we use the end-points method where the maximum value is assigned as 1 and the minimum value is assigned as 0. Another challenge is to map the so called "excellent", "good" and "poor" type of alternatives onto the scale. As these alter-

natives are some-what "fuzzy", we need to carefully assign them onto the interval scale. For example, "excellent" can be mapped onto values greater than 0.8250 and poor can be mapped onto values between 0 and 0.1250. This depends on the type of problem we are considering and is upto stakeholders and experimenters to decide. Once these are mapped onto numerical values, one can easily perform mathematical operations such as addition and multiplication to determine the alternatives [5]. As can be seen from figure 2, we have assigned highest value 1 to "excellent" and lowest value of 0 to "poor".

Once we have decided the intervals on the scale, we can linearly transform the scale such that highest number gets 1 and lowest gets 0. To calculate the expected utility of individual context states, we use the values on the scale as the utilities. For example, $U(S_{SA}^t = $ "excellent"$) = 1$ and $U(S_{SA}^t = $ "poor"$) = 0$. Similarly, "very good", "good" and "fair" are assigned utilities as 0.75, 0.50 and 0.25, respectively. Using e.q.1, stakeholders can determine the EU of each context state ($S_{QoE\_state}^t$). EU is then mapped on the scale and QoE is determined based on the closest matching value of the label on the bi-polar interval scale. For example, if $U(S_{SA}^t) = 0.8995$, it means that context state $S_{SA}^t = $ "excellent".

Once we have determined individual context states, these states are then fused together to determine the overall QoE (situation) of the user. It is denoted as R_{QoE}^t. For this, we consider a global utility which comprises of several QoE related states such as cognitive (user satisfaction in terms of the MOS) and behavioural (technology acceptance). There can be n number of QoE related classes which should be combined on a single scale to determine the overall QoE perceived by the user(s) [2]. To solve this problem, we define a global utility which is the linear weighted sum of the multiple QoE states and is calculated as follows:

$$U(R_{QoE}^t) = \sum_i^n w_i * A_i \qquad (2)$$

where $A_i = EU(QoE_i^n)$ is calculated using e.q. 1, w_i are the weights associated with individual QoE classes and states and $\sum_{i=1}^n w_n = 1$. A decision alternative, a is selected which matches the users interests on the bi-polar scale. Based on the mapped R_{QoE}^t, we can easily measure whether the overall QoE is "poor", "fair", "good", "very good" or "excellent".

3. IMPLEMENTATION AND EVALUATION

We developed a prototype using the GeNIe/SMILE APIs [6] to validate our proposed architecture. For results evaluation, we considered a VoIP application. In this section, we derive relationships between two different QoE classes: user satisfaction (S_{SA}^t) and technology acceptance (S_{TA}^t) and several context parameters to determine the overall QoE. We consider the ITU-T E-Model as a plug-in to determine S_{SA}^t. For S_{TA}^t,

```
Begin
Initialize:
1.   Define context attributes $(a_n^t)$ and QoE classes/states
$(S_{QoE\_state}^t)$.
2. Select the bi-polar scale based on application/experiment type.
Run:
Pass 1: QoE measurement
1. Create a Bayesian network (BN).
2. Learn the parameters of the BN using the initial training set.
3. Infer the beliefs of BN and determine hypotheses $(h_n)$ for each
$S_{QoE\_state}^t$.
4.  Calculate the expected utility $EU(S_{QoE\_state}^t)$ for each QoE
state $S_{QoE\_state}^t$.
5. Map the $EU(S_{QoE\_state}^t)$ onto the bi-polar scale and determine
QoE for that $S_{QoE\_state}^t$.
6. Calculate the overall QoE based a global utility $(U(R_{QoE}^t))$.
7. Map the $U(R_{QoE}^t)$ on the bi-polar scale and determine the over-
all QoE situation $(R_{QoE}^t)$.
8.  Record $R_{QoE}^t$ and append it to initial training set and call it
$trainingset_{append}$
Pass 2: QoE prediction
1. Perform linear-scale transformation of $U(R_{QoE}^t)$ in the $1^{st}$ BN
and convert it into the chance node.
2. Create a $2^{nd}$ BN including a new QoE chance node.
3. Learn the parameters of the $2^{nd}$ BN using $trainingset_{append}$
dataset.
4. Infer and predict $R_{QoE}^t$.
End
```

Fig. 3. Algorithm to determine and predict QoE.

we set the parameter ranges in relation to S_{SA}^t and location.
Here, we use the intuition that if a user is satisfied, he/she is
more likely to accept the technology/application/service. Ta-
ble 1 shows the values and ranges for several context param-
eters. The range for packet loss and delay were derived from
the current ITU-T G.107 and ITU-T G.113 recommendations
and experiments for ITU-T G.711 and ITU-T G.729 codecs.
We note that that ITU-T G.711 codec performs better at 0%
packet loss than the ITU-T G.729 codec but overall, ITU-T
G.729 codec performs significantly better even in case of low
to medium packet loss and delay. This is also true for the
cases when medium to high packet loss and delay were expe-
rienced by the test subjects. For both codecs, MOS degrades

Table 1. Selected context attributes and ranges.

Context attributes	values (range)
codec	ITU-T G.711, ITU-T G.729
packet loss (%)	low (0%-5%), medium (5%-7%), high (7%-10%)
delay (ms)	low (0 ms-150 ms), medium(150 ms - 300 ms), high (>300 ms)
location	home, train station, office

rapidly with an increase in delay.

Based on these observations, we set the conditional prob-
ability tables (CPTs). We further manipulated the CPTs to
include the effects of location on the overall QoE based on
the fact that users' social context changes at different loca-
tions. For example, we expect a user to be more comfortable
at home rather than his/her office due to stress or background
noise which might affect his/her QoE. This is reflected by as-
signing higher probability to home followed by tram stop and
the office to maximize R_{QoE}^t. Context states S_{SA}^t and S_{TA}^t
are inferred and then fused together to determine overall QoE
(R_{QoE}^t) as shown in the algorithm (figure 3). By using exact
algorithm such as clustering [4], we infer context states taking
values between 1 and 5. These values can help to understand
the hypothesis such as *"user satisfaction is fair"* or *"he/she
is not willing to accept the technology"*. In this case, it can be
derived by setting the evidence variables such as P(user sat-
isfaction="excellent" | location="home" delay="low" packet
loss= "low"). For the scope of this paper, we present eight
cases from a total of fourteen which show the capabilities of
our proposed approach. Our first goal was to perform con-
sistency checks to determine whether the proposed approach
was able to infer R_{QoE}^t correctly.

We conducted several experiments as shown in table 2.
Consider experiment 1, where a user is standing at the tram
stop during peak morning hours. He receives a call from
one of his colleagues. The mobile node calculates the ob-
served QoS related context parameters at this point in time.
Low packet losses and low delay were observed. To calculate
the R_{QoE}^t , S_{SA}^t and S_{TA}^t were inferred. Our algorithm pre-
dicts "very good" and "excellent" for S_{SA}^t with probabilities
0.80 and 0.20. In case of S_{TA}^t, the algorithm predicted "very
good", "good" and "fair" with probabilities 0.25, 0.50 and
0.25, respectively. In order to determine R_{QoE}^t , we make use
of e.q.1 and e.q. 2 where for each decision alternative, utili-
ties can be calculated and then mapped onto the bi-polar scale.
We assigned equal weights of 0.50 to each context state. The
overall utility calculated was 0.7375 and based on the scale
shown in figure 2, R_{QoE}^t is inferred as "very good".

Consider experiment 2, a user uses a VoIP application at
the tram stop and the observed packet loss and delay were
high. In this case, the overall QoE was "poor" with value
0 which was correctly determined by the model. Consider
experiment 3, where a user is sitting at his home and is re-
laxed. The observed packet loss and delay were low while
using ITU-T G.729 codec. In this case the R_{QoE}^t was inferred
as "very good". Similarly, we conducted several other experi-
ments as shown in table 2 to evaluate our proposed approach.

3.1. Quality of Experience Learning using Expectation Maximization Algorithm

In order to deal with uncertainty and missing information, we
use the EM algorithm [4] for learning BN parameters. We

Table 2. Observed context parameters with probabilities. Final QoE utility value U(QoE) is calculated and is mapped onto the bi-polar scale to determine R_{QoE}^t.

Exp.No.	codec	packet loss (%)	delay (ms)	location	U(QoE)	R_{QoE}^t
1.	ITU-T G.711	low (1)	low (1)	tram stop (1)	0.7375	very good
2.	ITU-T G.711	high (1)	high (1)	tram stop (1)	0	poor
3.	ITU-T G.729	low (0.90), medium (0.10), high (0)	low (0.60), medium (0.25), high (0.15)	home (1)	0.6768	very good
4.	ITU-T G.729	medium (1)	medium (1)	home (1)	0.50	good
5.	ITU-T G.729	high (1)	high (1)	tram stop (1)	0.0062	poor
6.	ITU-T G.711	low (1)	low (1)	home (1)	0.9750	excellent
7.	ITU-T G.711	low (0.20), medium (0.60), high (0.20)	low (0.30), medium (0.30), high (0.40)	tram stop (1)	0.1847	fair
8.	ITU-T G.729	low (0.80), medium (0.20), high (0)	low (0.40), medium (0.30), high (0.30)	office (1)	0.5616	good

generated several log files based on the ITU-T recommendations to emulate the feedback provided by users. The inputs were a_{delay}^t, a_{codec}^t, $a_{packetloss}^t$, $a_{location}^t$, S_{SA}^t and S_{TA}^t. After S_{SA}^t and S_{TA}^t are inferred, we determine the R_{QoE}^t which is then added to the initial training set. To predict R_{QoE}^t from the new observations, we create a new BN and re-train the model parameters using EM algorithm. We consider 14 test cases for QoE measurement and prediction. For each test case, we generated a log of 30 data items for 30 test subjects giving ratings for each test case. Thus, we had a total of 420 (30*14) data items in our training set. Based on this, we performed 10-fold cross validation and our model achieved classification accuracy of 94%. Thus, our results validate that our model is highly accurate and is able to make correct predictions.

Discussion: Based on our results, we show that our approach can efficiently measure and predict QoE based on two QoE states and several context parameters. Our approach is flexible as it allows the addition and deletion of QoE states and context attributes as and when required. The proposed approach ensures that QoE states can be measured correctly under uncertainty using BN, utility theory and the bi-polar scale. This is achieved by fusing these states probabilistically and by assigning appropriate utilities them. This helps in calculating the overall global utility which further determines the overall QoE. Our model can efficiently learn the probabilities on-the-fly and can predict correct QoE values. We are currently investigating the bounds on learning in the case of sparse data.

4. CONCLUSION AND FUTURE WORK

In this paper, we present a novel context-aware approach for quality of experience (QoE) measurement and prediction based on Bayesian networks, utility theory and context spaces theory (CST). Our proposed approach can incorporate several QoE classes and context parameters to correctly determine

and predict a single QoE value. We use the expectation maximization algorithm that can deal with learning under uncertainty and missing user data. We believe our approach can benefit network operators, codec engineers or application designers who are interested in measuring users' QoE in realistic settings. In future, we will extend this approach to incorporate sequential decision making in heterogeneous access networks.

5. REFERENCES

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

[2] P. Brooks and B. Hestnes, "User measures of quality of experience: why being objective and quantitative is important," *Network, IEEE*, vol. 24, no. 2, pp. 8–13, march-april 2010.

[3] A. Padovitz, S. Loke and A. Zaslavsky, "An approach to data fusion for context awareness," in *Fifth International Conference on Modelling and Using Context, CONTEXT'05*. 2005, pp. 353–367, Springer.

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

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

[6] Genie/smile software package and apis, http://genie.sis.pitt.edu/about.html, [online] access date: 01/02/11.