

Aplicación de Métodos No Intrusivos basados en BPNN (Backpropagation Artificial Neural Network) y RNN (Random Neural Network) para la evaluación de la Calidad de la Experiencia (QoE)

Diego J.L Botia Valderrama
Docente- Investigador
Grupo de Investigación GITA
Universidad de Antioquia
Medellin – Colombia
Email: diego.botia@gmail.com

Natalia Gaviria Gómez
Docente - Investigador
Grupo de Investigación GITA
Universidad de Antioquia
Medellin - Colombia
Email: ngaviri@gmail.com

Resumen

La medición y evaluación de la calidad de la experiencia (QoE), se ha convertido en uno de los principales puntos de atención en los operadores de telecomunicaciones para proveer servicios con la calidad esperada para sus usuarios, pero debido a la gran cantidad de factores que afectan la calidad del video, como los parámetros de red y de codificación empleados y la dificultad en la correlación entre las métricas objetivas y subjetivas, se hace muy complejo poder estimar la verdadera calidad percibida por los usuarios. En este trabajo se utilizó un modelo basado en redes neuronales artificiales, en el cual se evaluaron las redes BPNN (Backpropagation Neural Network) y las redes Random Neural Networks (RNN) empleando las métricas de calidad subjetiva MOS (Mean Opinion Score) y de calidad subjetiva PSNR (Peak Signal Noise Ratio), SSIM (Structural Similarity Index Metric), VQM (Video Quality Metric), QIBF (Quality Index Based Frame), definiendo una estrategia de QoS basada en Diffserv. Las métricas fueron evaluadas y analizadas mediante los coeficientes de correlación de Pearson, Spearman, RMSE y las tazas de outliers. Se obtuvieron valores de correlación superiores al 90% para todas las métricas evaluadas.

Palabras Claves: BPNN, RNN, Diffserv, WRED, QoS, QoE

I. INTRODUCCIÓN

En la implementación de plataformas de Televisión Digital (e.g. IPTV y DVB), siempre se presentan algunas restricciones clave que afectan la gestión y operación adecuada de la red, por ejemplo la gran cantidad de ancho de banda que debe contratar el usuario y que incrementa los costos del servicio, la limitante de los buffers internos en los routers y Set Top Box, que pueden generar problemas como pérdida de paquetes que son críticos en la transmisión de video o audio, el tipo de formato de compresión de video que permitirá reducir el uso del canal sin afectar la calidad y otros elementos que deben instalarse y configurarse adecuadamente como tipo de enlace en la última milla, tipo de control de admisión utilizado, entre otros.

Diferentes investigaciones han generado modelos de evaluación de la calidad sobre el video y el desarrollo de estrategias de medición para identificar los valores óptimos de las métricas, para garantizar la experiencia del televidente, por ejemplo existen proyectos como los nombrados por Winkler [2] VQEG (Video Quality Experts Group), QoSM (Quality of Service Metrics) del ATIS IPTV Interoperability Forum, y métricas específicas como las orientadas a paquetes, a bitstream, las híbridas y las métricas de imágenes [3] pero estas son cada vez más complejas, y no se han aplicado aún métodos de correlación para evaluación de servicios de tiempo real.

Uno de los principales problemas en la estimación de la calidad subjetiva del video (Métrica MOS), es la falta de modelos de estimación o correlación adecuados, que permitan garantizar resultados con gran exactitud y que además dependan fuertemente de las métricas objetivas, tanto FR (Referencia Completa) como RR (Referencia Reducida). Según Ries et.al [4] hay tres formas para realizar la estimación de la calidad del video. La primera es aplicar una evaluación subjetiva con un grupo de personas seleccionadas, cuyo principal inconveniente son sus costos y tiempo consumido. La segunda es la evaluación de las métricas de calidad objetivas de video, donde el principal problema es la baja correlación de estas métricas con respecto a la métrica de calidad subjetiva [5]; además, no tienen en cuenta los parámetros de red y de contenido. La tercera es utilizar métodos de aprendizaje de máquina para analizar la evaluación objetiva y subjetiva.

Los métodos no intrusivos de estimación de QoE para el video principalmente están basados en parámetros a nivel de aplicación o de red. Los modelos basados en redes neuronales artificiales han sido poco explorados, para la estimación y predicción de la calidad del video. Para la evaluación de los métodos no intrusivos, se decidió emplear 2 metodologías de machine learning y así evaluar la salida de cada sistema MOS Estimado versus el MOS Esperado, calculado de las métricas VQM [16], SSIM [8] [14], PSNR [13] y QIBF [18]. En cada caso, se calculó el ajuste de la correlación determinado por el coeficiente de correlación de Pearson, Spearman, RMSE para estimar el error y la tasa de outliers. Las metodologías probadas fueron: BPNN feedforward, y Red Neuronal Aleatoria (RNN).

Este artículo presenta inicialmente las principales métricas de calidad tanto objetivas como subjetivas, y las metodologías de evaluación de QoE no intrusivas basadas en técnicas de machine learning con sus principales características y funcionamiento. Luego en la sección tres se explica el testbed desarrollado, en la sección cuatro se presenta los resultados obtenidos y su correspondiente análisis y luego se presentarán las principales conclusiones.

II. MARCO REFERENCIAL

1. CALIDAD DE LA EXPERIENCIA (QoE).

La QoE es el grado de satisfacción de los usuarios con cualquier tipo de servicio multimedia. Este concepto ha sido definido de diferentes maneras por varios autores. Li-yuan et.al [7] indica que la QoE involucra dos aspectos: La primera es el monitoreo de la experiencia del usuario en línea y la segunda es el control del servicio para asegurar que la QoS (Calidad del Servicio) pueda conocer ampliamente los requerimientos del usuario. La QoE es una extensión de la QoS, porque el primero provee información sobre la entrega de servicios desde el punto de vista del usuario

final. A partir de estas definiciones de *QoE*, se puede concluir que ésta hace referencia a preferencias personales de los usuarios, y por tanto busca evaluar la percepción subjetiva que tienen del servicio recibido. Sin embargo, esta percepción está influenciada en alto grado por el desempeño de la red en cuanto a los parámetros de *QoS*.

Según Mosheni [6], la evaluación de la calidad del video es un problema multidimensional que requiere múltiples criterios, algunas veces localizados en distintos niveles, los cuales se explicaran más adelante. Existen factores comerciales por parte de las telcos (Por ejemplo contenido, precio, usabilidad, etc), que pueden atraer nuevos suscriptores pero la *QoE* es un factor crítico para poderlos retener. De acuerdo a Winkler et.al [1] en la *QoE* existen dos tipos de medidas para evaluar la calidad del video transmitido sobre una red: las subjetivas y las objetivas. Las métricas subjetivas dependen de las expectativas de los usuarios y por lo tanto tiene en cuenta los sentimientos, percepciones y opiniones de ellos. Por consiguiente, los métodos subjetivos son utilizados para establecer el performance de un sistema de televisión usando mediciones que están más directamente relacionadas con la percepción de los usuarios. Para realizar las pruebas subjetivas se emplean en promedio entre 15 y 30 personas que observan una serie de video clips y los evalúan en un rango específico. Para ello se emplea la métrica *MOS* (*Mean Opinion Score*) que es la más conocida, y fue propuesta por ITU Ibid [8]. Esta es una simple medida numérica utilizada en tráfico multimedia, el cual determina la calidad percibida de flujos de video o audio que son recibidos después del proceso de codificación y /o transmisión. Esta medida esta en un rango que va del 1 (calidad muy baja) al 5 (calidad muy alta).

De acuerdo a Kuipers et.al[9], el mínimo umbral que se considera como una calidad aceptable corresponde a un *MOS* de 3.5. Aunque las pruebas subjetivas (como *MOS*) son muy útiles en la medición de la satisfacción del usuario y que además generan una verdadera evaluación de la subjetividad de la persona gracias a su exactitud, la aplicación de ellas es aún muy compleja debido al elevado consumo de tiempo, dinero, no son flexibles, son imprácticos para poderlos probar con equipos de red y además se requiere de un ambiente controlado que en algunos casos es difícil de implementar. Además, si se desea implementar técnicas de gestión de tráfico en tiempo real, es necesario encontrar una relación de éstas con las métricas objetivas, medibles por los equipos de red

Es así, como los métodos objetivos están basados en algoritmos para la evaluación de la calidad de video, que los hace menos complejos, además se pueden realizar sobre ambientes de simulación controlados. Las métricas objetivas se soportan en modelos matemáticos que se aproximan al comportamiento del Sistema de Visión Humana (*SVH*) y por consiguiente, predecir lo más exactamente posible la verdadera *QoE*. Sin embargo la percepción de cada persona es altamente influenciada por la calidad de la red de datos expresada por la definición de los parámetros de *QoS*. Existe una gran cantidad de propuestas para evaluación de métricas *QoE* y así determinar la verdadera experiencia del usuario. La ITU ha realizado un trabajo de estandarización de algunas de ellas [10][11].

A pesar de todas las propuestas de métricas objetivas, aún existe el problema de que no siempre se correlacionan de manera adecuada con la percepción humana, porque esta percepción esta altamente

influenciada por el performance de la red, debido a los parámetros de *QoS* definidos. De acuerdo con [12] las métricas objetivas son modelos computacionales que predicen la calidad de la imagen percibida por una persona y pueden ser clasificadas en métodos intrusivos y no intrusivos, mostrados en la figura 1.



Figura 1. Métodos de Evaluación de calidad de la experiencia. Fuente El autor

Uno de los principales problemas en la estimación de la calidad subjetiva del video (Métrica *MOS*) es la falta de modelos de estimación o correlación adecuados, que permitan garantizar resultados con gran exactitud y que además dependan fuertemente de las métricas objetivas, tanto *FR* (Referencia Completa) como *RR* (Referencia Reducida)[13][1]. Debido a factores complejos como el *HVS*, se han propuesto diferentes tipos de soluciones, como las aplicaciones con técnicas de aprendizaje de máquina (*machine learning*). Algunos de los métodos más utilizados son: Redes neuronales artificiales (*ANN*), lógica difusa, redes neuro – difusas (por ejemplo *ANFIS*), máquinas de soporte vectorial (*SVM*), procesos gaussianos, entre otros. Los métodos no intrusivos de estimación de *QoE* para el video principalmente están basados en parámetros a nivel de aplicación o de red. Los modelos basados en redes neuronales artificiales han sido poco explorados, para la estimación y predicción de la calidad del video.

Como se observa en la figura 1, existen técnicas no intrusivas, que son clasificadas en *RR* y *NR*. El modelo de evaluación de la calidad Pseudo Subjetivo o *PSQA* es un ejemplo de esta última categoría. Este modelo utiliza una red *RNN* (*Random Neural Network*) para aprender y reconocer la relación entre el video y las características de la red con la calidad percibida por los usuarios. Inicialmente, para realizar el proceso de entrenamiento de la *RNN*, se debe tener una base de datos que contenga diferentes secuencias para evaluar las distorsiones generadas por varios parámetros de *QoS* y codificación. Después del entrenamiento de la *RNN*, se evalúa cualquier tipo de secuencia (P.ej tiempo real), y por consiguiente, el *MOS* es validado.

Basado en el trabajo de Zinner et.al[15], el cual presenta un framework para la evaluación de la *QoE* utilizando sistemas de video streaming. Botia et.al [17] propone un nuevo mapeo entre las métricas *PSNR*, *SSIM*, *VQM* y *MOS* relacionado en la tabla 1.

III. TESTBED DE LA SIMULACION

En la simulación se utilizó una selección de diferentes secuencias de video *Raw* sin comprimir en formato *YUV* con modo de color de video o *sampling* de 4:2:0, codificados con las herramientas de software *ffmpeg* y *main concept*, para adaptarlos a diferentes tasas de bits y longitudes *GOP*. Se evaluaron inicialmente 4 secuencias de video con diferentes niveles de movimiento, codificados en el formato *MPEG-4* que fueron adaptadas para poderlas transmitir por una red IP simulada.

Tabla 1. Mapeo QoE con las Métricas SSIM, PSNR, VQM y MOS [17]- [18]

MOS	PSNR (dB)	SSIM	VQM	QIBF
5(Excelente)	≥ 37	$\geq 0,93$	$< 1,1$	$\geq 0,85$
4 (Bueno)	$\geq 31 - < 37$	$\geq 0,85 - < 0,93$	$\geq 1,1 - < 3,9$	$\geq 0,65 - < 0,85$
3(Aceptable)	$\geq 25 - < 31$	$\geq 0,75 - < 0,85$	$\geq 3,9 - < 6,5$	$\geq 0,45 - < 0,65$
2 (Pobre)	$\geq 20 - < 25$	$\geq 0,55 - < 0,75$	$\geq 6,5 - < 9$	$\geq 0,25 - < 0,45$
1 (Malo)	< 20	$< 0,55$	≥ 9	$< 0,25$

La figura 2 presenta algunas capturas de los videos evaluados (Noticias de la TV pública española, Misa, Highway y Árbol en invierno)¹, convertidas y codificadas a una resolución de 720x480 pixeles (definición estándar) bajo el estándar NTSC, y con *frame rate* de 30 fps. Por cada flujo de video varios parámetros fueron combinados como la longitud del *GOP* (10,15 y 30), las tasas de bits recomendadas por *DSL Forum* [32] (1.5, 2, 2.5 y 3 Mbps) y tasa de pérdida de paquetes para ambas redes *Besteffort* y *Diffserv* (empleando el algoritmo de control de congestión *WRED*), lo cual produjo 385 diferentes secuencias de video para probar [17].

Las trazas de video generadas, se adecuaron para enviarlas a la red de datos a través de la encapsulación de cada paquete con un *MTU* (*Maximum Transfer Unit*) de 1024 bits, utilizando el protocolo *RTP* (*Real Time Transport Protocol*) por medio de la herramienta *MP4trace*. Mediante *NS-2* y el *framework Evalvid*[19] se generaron los archivos *sender trace files* y *receive trace files*, que permitieron calcular los *frames* y paquetes enviados, recibidos y perdidos, los retardos y los *jitters*, que facilitaron el análisis de cada secuencia de video para ambos escenarios implementados. El *Framework Evalvid*[19] además soporta las métricas *PSNR* y *MOS* y posee una estructura modular, que lo hace fácilmente adaptable a cualquier ambiente de simulación. Se utilizó además la herramienta *MSU VQMT* [31] que permite obtener las métricas *Y-PSNR*, *SSIM* y *VQM* a través del video de referencia original y el video con distorsión recibido.

El escenario de simulación, esta formado por un *video sender* (servidor de video bajo demanda) y 9 fuentes de tráfico cruzado, que consisten en fuentes de tráfico *CBR* y *On-Off*. Así se permitirá realizar el envío de los paquetes de video sobre una red con congestión y permitirá probar los esquemas de *QoS* definidos. El flujo de video *MPEG-4* se completa con flujos de tráfico *background* de tipo *On-Off*, que tiene una distribución exponencial con un tamaño de paquete promedio de 1500 *bytes*, tiempo de ráfaga (*burst time*) de 50ms, tiempo de inactividad (*idle time*) de 0.01 *ms* y tasa de envío de 1 *Mbps* [17][18]. La red de acceso esta representada por un video receiver (simulando una última milla con *ADSL2*) con un enlace de 10*Mbps* de ancho de banda y varios nodos de recepción (*sink*) para el tráfico cruzado con un ancho de banda de 10*Mbps* por cada uno. Las distorsiones de la transmisión fueron simuladas a diferentes tasas de pérdidas de paquetes (*PLR*). Se probó el comportamiento del tráfico y de las métricas QoE con diferentes porcentajes de error, sobre un enlace establecido entre el router core y el router de borde, utilizando un modelo de pérdidas con distribución uniforme con tasas de 0%, 1%, 5% y 10% y retardo de 5 ms.



Figura 2. Capturas de las secuencias de video evaluadas. Fuente El autor

IV. APLICACIÓN DE MÉTODOS NO INTRUSIVOS PARA ESTIMAR LA CALIDAD DEL VIDEO MEDIANTE MÉTRICAS OBJETIVAS Y SUBJETIVAS.

Para la evaluación de los métodos no intrusivos, se decidió emplear dos metodologías de *machine learning* (*BPNN feedforward* y Red Neuronal Aleatoria - *RNN*); y así obtener la salida de cada sistema para evaluar el *MOS* Estimado versus el *MOS* Esperado; este ultimo calculado por medio del mapeo de las métricas *VQM*, *SSIM*, *PSNR* y *QIBF*. En cada caso, se calculó el ajuste de la correlación determinado por el coeficiente de correlación de *Pearson* y el *RMSE* para estimar el error. En la figura 3, se presenta la metodología general empleada [20].

Se utilizó el software *Matlab* para *BPNN*, empleando el toolbox *Neural Networks*. Para el caso de las redes *RNN*, se empleo el software *QoE-RNN* [21]. De las 385 secuencias distintas con diferentes parámetros configurados, se emplearon el 70% como datos de entrenamiento y el 30% restante como datos de validación y prueba. A continuación, se mostrará y se discutirá los resultados encontrados en cada máquina de aprendizaje seleccionado.

2.1 Caso 1. Aplicación de una Red Neuronal Artificial Feed Forward con Backpropagation

Las *ANN* tipo *feed – forward* son las más empleadas para realizar estimación sobre las métricas subjetivas. Varios trabajos proponen diferentes modelos basados en *ANN*, [23], [24], [25]. Estas aproximaciones, generalmente se aplican sobre sistemas móviles, o con secuencias de video con bajas resoluciones (*QCIF* o *CIF*), además no tienen en cuenta parámetros de red o métricas objetivas de video. En la gran mayoría de casos, se emplean una o dos métricas como entradas a la red y los resultados de las autocorrelaciones de *Pearson* casi siempre se ubican por debajo de 0.90. Según Ding et.al[22], la red neuronal puede ser utilizada para obtener funciones mapeadas entre los índices de evaluación de calidad objetivo y la evaluación de calidad subjetivo. Esta afirmación permite comprender la utilidad de las *ANN* para analizar el *MOS* estimado y del modelo propuesto presentado en este trabajo.

¹ Secuencias cedidas gracias al grupo G@TV de la Escuela Técnica Superior de Ingenieros de Telecomunicación E.T.S.I.T. Universidad Politécnica de Madrid (España).

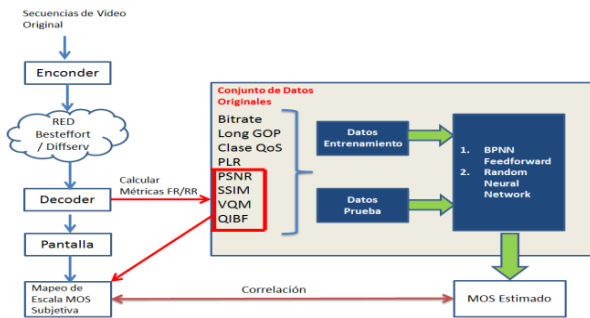


Figura 3. Metodología propuesta para estimar el MOS mediante técnicas de Machine Learning. Fuente El autor.

En este caso, se realizará el entrenamiento de la red a través de los parámetros descritos anteriormente, que se convertirán en las variables de entrada. Como se ha indicado, los parámetros objetivos pueden llegar a afectar la calidad del video. Después del entrenamiento, se procederá a realizar una evaluación con un conjunto de datos de prueba (96 secuencias) para realizar la respectiva validación de la red. La idea es llegar al error más bajo y poder correlacionar el MOS estimado por la ANN versus el MOS Promedio calculado de las métricas objetivas definidas por la tabla 1.

Las ANN son un paradigma para el procesamiento de la información que es inspirado en el sistema neuronal del ser humano. Generalmente, las ANN están compuestas de un gran número de elementos de procesamiento altamente interconectados llamados neuronas, que trabajan en conjunto para resolver problemas [4]. La base es la creación de una red neuronal llamada también perceptrón multicapa (MLP) generalmente dividida en 3 o N-capas. La primera capa son las neuronas conectadas a los datos de entrada, la segunda capa se conoce como la capa oculta y contiene un conjunto de sinapsis así como una serie de pesos W_{ij} y unas funciones de activación definidas para excitar o inhibir cada neurona, generando una respuesta. La tercera capa es la capa de salida conectada directamente con la capa oculta y en ella se obtendrán los vectores con los datos de cada salida estimada. Las múltiples capas de las neuronas con funciones de transferencia no lineal (Como por ejemplo Tangencial – sigmoidea) permite a la ANN aprender las relaciones lineales y no lineales entre las entradas (PLR, GOP, Bitrate, Clase de QoS, QIBF, PSNR, SSIM, VQM) y los vectores de salida deseada (MOS). En la figura 4, se presenta la estructura de la red neuronal artificial *backpropagation*.

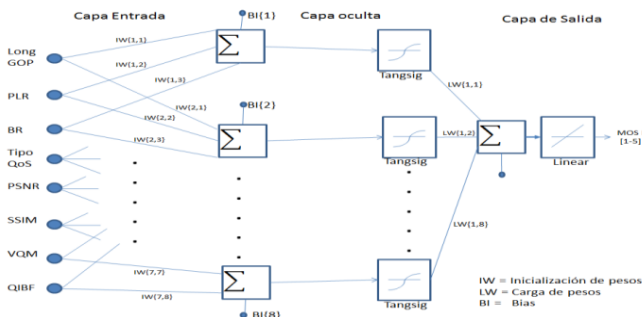


Figura 4. Arquitectura propuesta para estimación de MOS mediante una red BPNN Feed Forward. Fuente el Autor

Al igual que un ser humano, el sistema ANN necesita de una etapa de aprendizaje y otra de validación y prueba, para establecer si la red neuronal generalizó su aprendizaje para cualquier conjunto de datos.

Para este proceso, se entrena la red mediante algún algoritmo de entrenamiento y se calcula el error más bajo posible, mediante el cálculo de la función de costo MSE (Mean Square Error). Se utiliza el algoritmo iterativo del gradiente descendente u otros algoritmos de aprendizaje para lograr la convergencia a un valor objetivo (target), es decir, para calcular el mínimo error de entrenamiento de la red.

Para el caso de estudio se procede a construir la función MOS_Estimado, definida por la ecuación 1

$$MOS_Est = f(x_1, x_2, \dots, x_n) \quad (1)$$

donde x_1, x_2, \dots, x_n son los parámetros de entrada establecidos por el bitrate, packet loss rate, longitud GOP, Clase de QoS (1 para Besteffort y 2 para Diffserv), SSIM, PSNR, VQM y QIBF.

En la figura 5, se presenta la arquitectura propuesta de la Red Neuronal BPNN feed-forward. En la capa de entrada y la capa de salida, las neuronas tiene una función de activación lineal (purelin), en la capa oculta después de hacer varias pruebas se obtuvo el menor error de entrenamiento con 20 neuronas y con función de activación de tipo tangente -sigmoidea (tansig). Para alcanzar la función objetivo se empleo el algoritmo de entrenamiento Levenberg-Marquardt.

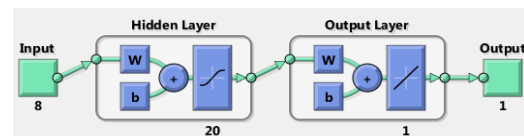


Figura 5. Arquitectura de la Red Neuronal Feedforward de 3 capas. Fuente El Autor

A continuación se procedió a realizar la etapa de validación, en donde se obtuvo los resultados de la figura 6.

Como se observa se obtuvo un buen ajuste entre los datos de salida de la red y los datos del MOS deseado. Se realizó un análisis de la regresión lineal entre ellos. La relación se establece entre el valor estimado por la red BPNN (variable y) y los datos deseados (x). La representación esta dada por la ecuación lineal clásica (Ecuación 2):

$$y = m \cdot x + b \quad (2)$$

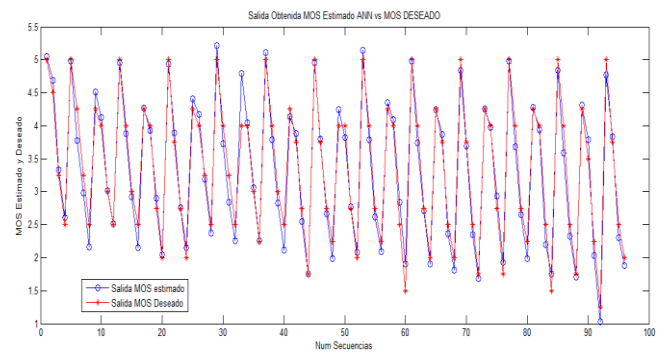


Figura 6. Resultado de la salida obtenida – MOS Estimado vs MOS Deseado. Fuente : El autor

Según el parámetro de correlación de Pearson entre el MOS estimado por la red BPNN y el MOS deseado, se obtuvo un ajuste lineal de 96.72% y un RMSE de 0.1977. La figura 7 presenta la salida de la correlación obtenida.

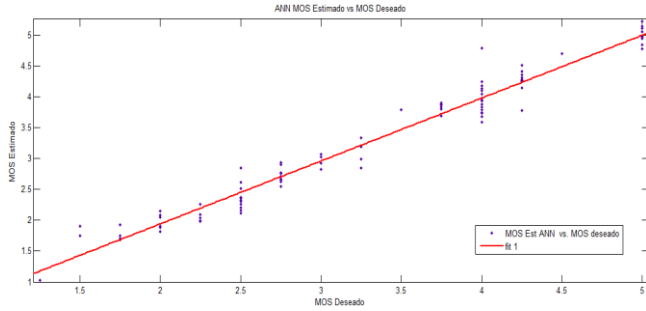


Figura 7. ANN MOS estimado vs MOS Deseado. Fuente El Autor

Con los resultados obtenidos se establece que la Red neuronal *feed forward* permitió una buena generalización con los datos utilizados para la validación y se determina una completa relación lineal.

2.2. Aplicación de una Red Neuronal Aleatoria (RNN) por medio de la metodología PSQA.

Esta clase de redes capturan con gran exactitud y con robustez el mapeo de funciones, donde varios parámetros están involucrados. Según Casas et.al [26], las redes RNN se ha empleado en varios campos de la ingeniería entre los que se destacan la resolución de problemas de optimización NP completos (*Non-deterministic Polynomial time*), problemas de imágenes de generación de texturas, algoritmos de compresión de imagen y video y problemas de clasificación de calidad percibida en voz y video sobre IP; que las hacen ideales para su aplicación en la evaluación de la QoE.

Las redes RNN son una mezcla entre las redes neuronales y el encolamiento en redes. Por definición, las RNNs son conjuntos de redes ANN compuestas por una serie de neuronas interconectadas. Estas neuronas intercambian señales que viajan instantáneamente de una neurona a otra y envían señales desde y hacia el ambiente. Cada neurona esta asociada a una variable aleatoria entera asociada a un potencial. El potencial de una neurona i en un tiempo t se define por $q_i(t)$.

Si el potencial de la neurona i es positivo la neurona esta excitada y aleatoriamente envía señales a otras neuronas o al ambiente de acuerdo a un proceso de Poisson de tasa r_i . Las señales pueden ser positivas (+) o negativas (-). Así, la probabilidad que la señal se envíe de la neurona i a la neurona j , sea positiva, se denota por $P_{i,j}^+$, y la probabilidad que la señal sea negativa se denota por $P_{i,j}^-$. La probabilidad que la señal vaya al ambiente se denota por d_i . Si N es el número de neuronas, para todo $i = 1, \dots, N$, entonces d_i se expresa como se ve en la ecuación 3

$$d_i + \sum_{j=1}^N (p_{i,j}^+ + p_{i,j}^-) = 1 \quad (3)$$

Por consiguiente, cuando una neurona recibe una señal positiva desde otra neurona o desde el ambiente su potencial se incrementa en 1. Si se recibe una señal negativa su potencial decrece por 1. Cuando una neurona envía una señal positiva o negativa su potencial decrece en una unidad.

El flujo de señales positivas que arriban desde el ambiente a la neurona i es un proceso de Poisson con una tasa λ_i^+ o λ_i^- . Es así posible tener $\lambda_i^+ = 0$ y $\lambda_i^- = 0$ para alguna neurona i . Para tener una red activa se necesita entonces la ecuación 4:

$$\sum_{i=1}^N (\lambda_i^+) > 0 \quad (4)$$

Definiendo g_i como la probabilidad de equilibrio para una neurona i en estado de excitación, se considera la ecuación 5

$$g_i = \lim_{t \rightarrow \infty} Pr(q_i(t) > 0) \quad (5)$$

En esta ecuación, si para un proceso de Poisson de las neuronas, $\vec{q}(t) = \{q_1(t), \dots, q_N(t)\}$, es ergódico, se define la red como estable y satisface las condiciones para un sistema no lineal.

En las redes RNN, el objetivo del proceso de aprendizaje es obtener los valores de R_i y las probabilidades $P_{i,j}^+$ y $P_{i,j}^-$. Lo anterior permite obtener los pesos de las conexiones entre las neuronas i,j , como se muestran en la ecuación 6:

$$\omega_{i,j}^+ = R_i P_{i,j}^+ \quad \text{y} \quad \omega_{i,j}^- = R_i P_{i,j}^- \quad (6)$$

De la ecuación, el conjunto de pesos en la topología de red se inicializa con valores arbitrarios positivos y con K iteraciones que son realizadas para modificar los pesos. Para $k=1, \dots, K$, el conjunto de pesos para el paso k se calcula desde el conjunto de pesos en el paso $k-1$. Sea $R^{(k-1)}$ la red obtenida después del paso $k-1$ definido por los pesos $\omega_{i,j}^{+(k-1)}$ y $\omega_{i,j}^{-(k-1)}$, entonces el conjunto de tasas de entradas (señales externas positivas) en $R^{(k-1)}$ para $X_i^{(k)}$ permitirá obtener una red que permita generar una salida $\vec{Y}^{(k)}$ cuando la entrada es $\vec{X}^{(k)}$; por consiguiente, $\vec{Y}^{(k)} = MOSs$.

La red RNN tiene una arquitectura de tres capas. Así, el conjunto de neuronas $R \in \{1, \dots, N\}$, se parte en 3 subconjuntos. El conjunto de neuronas de entrada, el conjunto de neuronas ocultas y las neuronas de salida. Las neuronas de entrada reciben señales positivas desde el exterior. Para cada nodo i se tiene que $\lambda_i^+ > 0$ y $d_i = 0$. Para los nodos de salida se tiene que $\lambda_i^+ = 0$ y $d_i > 0$. Los nodos intermedios no están directamente conectados al ambiente, para cualquier neurona oculta i se tiene que: $\lambda_i^+ = \lambda_i^- = d_i = 0$.

Para el caso de estudio, se tienen varias secuencias de video con diferentes parámetros, de donde se seleccionan un conjunto de datos de entrenamiento y otro conjunto de datos de prueba. En las ecuaciones 7, 8 y 9, S representa el conjunto de secuencias de entrenamiento donde cada secuencia esta definido por α_n y S' refiere al conjunto de secuencias de validación. Por otra parte, P es el conjunto de parámetros λ que afectan cada secuencia, donde:

$$S = \{\alpha_1, \alpha_2, \dots, \alpha_S\} \quad (7)$$

$$S' = \{\alpha'_1, \alpha'_2, \dots, \alpha'_S\} \quad (8)$$

$$P = \{\lambda_1, \lambda_2, \dots, \lambda_p\} \quad (9)$$

El valor del parámetro λ_p en la secuencia α_s esta definido por V_{ps} , donde $V = (V_{ps})$; siendo s una matriz. $s=1, 2, \dots, S$. Cada secuencia α_s recibe un puntaje MOSs $\in [1, 5]$. Por otra parte, para las secuencias $\alpha'_s \in S'$, se obtiene una función $f(V_{1s}, V_{2s}, \dots, V_{ps}) \approx MOSs$ y el proceso de entrenamiento termina; si no es así se pueden probar más datos o cambiar algunos parámetros de la RNN y se procede a construir una nueva función f .

Según Casas et.al [26], es necesario aplicar una metodología para evaluar la calidad de la experiencia basada en el uso de parámetros de red (probabilidad de pérdida de paquetes, retardos, jitters, entre otras) y parámetros de video (codificación, bitrate, frame rate, longitud de GOP, entre otras), que se convertirán en parámetros de entrada. Basado en estos criterios, se puede generar una función de mapeo

entre estos parámetros y el valor de calidad subjetivo definido por la métrica *MOS*. Para realizar esta tarea, proponen la metodología *PSQA* (*Pseudo Subjective Quality Assessment*), que utiliza las redes *RNN* para aprender el mapeo entre los parámetros y la calidad percibida [27][28]. Esta metodología tiene como características su exactitud, permite generar evaluaciones automáticas y en tiempo real, es eficiente y puede ser aplicado en varios tipos de media códec, y bajo diferentes parámetros y condiciones de red. Además, se puede extender para la comparación con métricas objetivas, que permita generar correlaciones mucho más precisas.

La red *RNN* se considera como una máquina de aprendizaje supervisada, que utiliza un par de características de red y multimedia y los valores *MOS*. Si en la etapa de entrenamiento se encuentra una relación de los parámetros de entrada a través de las métricas objetivas y la salida esperada, será posible estimar y/o predecir los valores subjetivos con mayor nivel de precisión. Debido a las características de las redes *RNN*, se podrían considerar ideales para dar buenas evaluaciones, cuando se produce una amplia variación sobre todos los parámetros que afectan la calidad [28]. Por esta razón, es un modelo exacto, rápido y de bajo costo computacional.

Para el desarrollo del caso de estudio propuesto, se procedió a implementar una arquitectura de red *RNN* feed-forward de 3 capas, propuesta por Mohamed et.al [27]. Se empleó el software *QoE-RNN*, que permite estimar el valor *MOS* a través del empleo de una red *RNN*. Esta herramienta posee licencia *LGPL* y esta desarrollada en el lenguaje de programación *C*.

En la figura 8, se presenta el proceso general donde se evalúan las secuencias de video transmitidas por la red, dependiendo de la estrategia de *QoS* implementado, se establecen los valores *MOS* relacionados a cada métrica objetiva (*PSNR*, *VQM*, *SSIM* y *QIBF*), y luego se calcula un *MOS* promedio ($MOS_p = \frac{\sum_{i=1}^N MOS_{obj_i}}{N}$) que será el parámetro objetivo de la red *RNN*. Aparte, se elijen los parámetros de entrada y con el MOS_{psqa} obtenido de la red, se procede a realizar el respectivo análisis de correlación.

El valor del número de iteraciones límite fue colocado en 2000 y la topología de la red, en este caso, fue de 9 neuronas de entrada, 10 neuronas en la capa oculta, y 1 neurona de salida. En las diferentes pruebas realizadas el mejor ajuste se obtuvo con 10 neuronas en la capa oculta.

Mediante *Matlab*, se realizó el análisis de correlación lineal entre el MOS_{psqa} y el *MOS* esperado. Se observa un buen ajuste lineal alcanzando un coeficiente R^2 de 0.9812 y un *RMSE* de 0.1412. En la figura 9, se presenta esta correlación.

En la tabla 2 se presenta el resumen general de todas las correlaciones obtenidas para cada caso de estudio.

El performance de las métricas de calidad perceptual depende de su correlación con los resultados objetivos.

Es así como la exactitud en las estimaciones de las métricas subjetivas se evalúan con respecto a la exactitud de la predicción, la monotonicidad y la consistencia en relación con la evaluación subjetiva de la calidad del video sobre un rango de secuencias de video de prueba con diferentes artefactos.

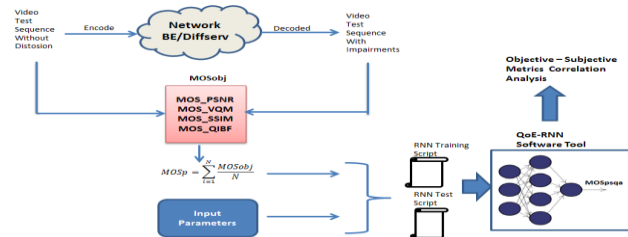


Figura 8. Metodología empleada para la aplicación de PSQA con la red RNN para el caso de estudio. Fuente: El autor

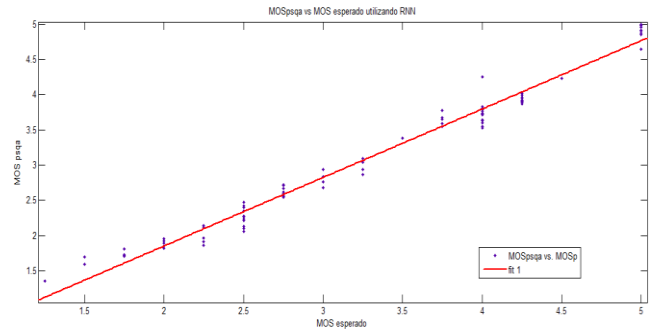


Figura 9. Correlación entre el MOS_{psqa} y el *MOS* esperado por medio de la red RNN. Fuente: El autor

Tabla 2. Resumen General de las correlaciones obtenidas para los Casos de Estudio

Correlación	PCC	SROCC	RMSE	OR
Y-PSNR vs MOS BPNN	0.9303	0.9713	0.2882	0,5833
VQM vs MOS BPNN	0.9412	0.9707	0.2647	0,1979
SSIM vs MOS BPNN	0.9446	0.9733	0.257	0,1483
QIBF vs MOS BPNN	0.937	0.9711	0.2739	0,1562
Y-PSNR vs MOS psqa	0,9698	0.9873	0.1796	0,5833
VQM vs MOS psqa	0.9645	0.9900	0.1948	0,1666
SSIM vs MOS psqa	0.9497	0.9889	0.2319	0,1354
QIBF vs MOS psqa	0.9254	0.9853	0.2824	0,1354

Los métodos de evaluación del performance son propuestos por el *Video Quality Experts Group (VQEG)* [29]. Existen 4 medidas estadísticas que se utilizan para evaluar el performance de las métricas de calidad de video: el coeficiente de correlación de *Pearson (PCC)*, el coeficiente de correlación de orden de rangos de *Spearman (SROCC)*, el error cuadrático medio (*RMSE*) y el *outlier ratio (OR)*. [30]

Como se observa en los resultados de la tabla 2, las correlaciones entre las métricas MOS_{psqa} fueron buenas con las métricas objetivas especialmente con *Y-PSNR* y *VQM*, aunque el coeficiente Spearman se mantuvo alto con todas las métricas lo que demuestra una alta tendencia lineal en todos los casos. La generalización fué muy alta y se pudo comprobar la consistencia, exactitud, y monotonicidad de los resultados.

En todos los casos, la correlación alcanzó valores superiores al 90%, y la correlación entre las métricas objetivas y subjetivas en cada caso presento un buen comportamiento lineal. Por consiguiente, las métricas como *VQM*, *SSIM* y *QIBF* se relacionan en gran medida con la subjetividad establecida por *MOS*. También, se observa que los resultados de la red *BPNN feed forward* presentaron una buena generalización del *MOS* estimado contra cada métrica objetiva evaluada.

V. CONCLUSIONES

En este trabajo, se obtuvo mejores valores de correlación entre las métricas de *QoE* objetivas y subjetivas gracias al empleo de los métodos no intrusivos y se validaron su exactitud, consistencia y monotonicidad a través del análisis de las correlaciones de *Pearson*, *Spearman*, tasa de *Outliers* y la raíz del error cuadrático medio (*RMSE*). Uno de los principales problemas de los métodos objetivos y subjetivos, es la falta de metodologías completas para evaluar la *QoE* de forma exacta. Por consiguiente, se abordó este problema con el fin de proponer una nueva metodología que permitió encontrar nuevas correlaciones entre las métricas objetivas y subjetivas, además, se propuso la aplicación de técnicas de aprendizaje de máquina por medio de Redes neuronales artificiales *back-propagation* y redes neuronales aleatorias, para aproximar mejor a la estimación de la percepción humana. En las diferentes simulaciones realizadas, se observó que las métricas, *VQM*, *SSIM* y *QIBF* son altamente correlacionadas y son próximas a la percepción de los usuarios (determinado por la métrica *MOS*). A diferencia de los trabajos previos, se desarrolló un modelo de correlación general, que utilizó parámetros de red y codificación aplicados sobre varias secuencias de video. Analizando los resultados con las máquinas de aprendizaje; las redes *BPNN* y *RNN* generaron altas correlaciones con las métricas objetivas con valores de *PCC* superior al 90% y con bajas tasas de error. La consistencia de las correlaciones entre las métricas a través de los *Outliers* se determinó con valores bajos.

VI. BIBLIOGRAFIA

[1] Winkler,S, Mohandas. P. The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics. IEEE TRANS.BROADCASTING VOL. 54, NO. 3, SEPTEMBER 2008. [on line] Disponible en <http://www.pgea.unb.br/~mylene/PI/WinklerMohandas2008.pdf>. 2008

[2] Winkler. S. Standardizing Quality Measurement for Video Services. IEEE COMSOC MMTTC. Vol.4, No.9, October 2009.

[3] Sheikh. H.R, Bovik. A.C, "Image information and visual quality," IEEE Transactions on Image Processing, vol.15, no.2,pp. 430- 444, Feb. 2006.

[4] Ries,M, Kubanek,J, Markus.R. "Video Quality Estimation for mobile streaming applications with neural networks". [online]. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.130.318>. 2006.

[5] Nemethova, O. Ries, M, Siffel E, Rupp, M. Quality Assessment for H.264 Coded Low Rate and low resolution Video Sequences". Proc . of Conf. on Internet and Inf. Technologies (CIIT), pp 136-140. St. Thomas, US Virgin Islands, 2004.

[6] Mohseni, S. Driving Quality of Experience in mobile content value chain. 2010 4th IEEE International Conference on Digital Ecosystems and Technologies (DEST). pp 320 - 325. ISSN :2150-4938. 2010

[7] L. Li–yuan, Z. Wen–an, and S. Jun–de, "The Research of Quality of Experience Evaluation Method in Pervasive Computing Environment," 1st Int. Symp. on Pervasive Comp. and Appl, pp. 178– 182. 2006

[8] ITU FG-IPTV.DOC-0814. "Quality of Experience Requirements for IPTV Services". Diciembre. 2007.

[9] Kuipers,D, Kooij,R, Vleeshauwer,D, Brunnstrom,K, "Tech. for Measuring QoE". Book Series: Lecture Notes in Comp. Science Ed. Springer ISSN 0302-9743 Pag 216-227. [online] Available: <http://www.springerlink.com/content/5028804658914365>. Mayo 2010.

[10] ITU-T Recommendation P.910. "Subjective video quality assessment methods for multimedia applications". Int Telecom Union. Genova, Suiza, 1999.

[11] ITU-T Recommendation. QoE requirements in consideration of Service Billing for IPTV Service, UIT-T FG-IPTV. 2006.

[12] Casas. P, Belzarena.P, Irigaray. I , Guerra. D. A User Perspective for End-To-End Quality of Service Evaluation in Multimedia Networks. LANC '07, San José, Costa Rica. 2007

[13] Boavida.F, Cerqueira.E, Chodorek.R, Grega.M, Leszczuk.M, Papir.Z, Romaniak. R, Guerrero.C, "Benchmarking the Quality of Experience of Video Streaming and Multimedia Search Services: The content Network of

Excellence". KSTiT 2008- XXIII Symp of Telecom and Teleinf. Inst of Telecom and Elec of University of Tech and Life Science, Bydgoszcz, Poland, 10-12 Sept 2008.

[14] Wang.Z, Bovik.A, Sheikh.H, Simoncelli. E. "Image Quality Assessment: From Error Visibility to Structural Similarity". IEEE Trans on Image Process. Vol 13, Nro 4. Apr 2004.

[15] Zinner. T, Abboud. O, Hohlfeld. O, Hossfeld.T, Train-Gia. P. Towards QoE Managment For Scalable Video Streaming. In 21st ITC Specialist Seminar on Multimedia Applications Traffic, Performance and QoE. Marzo 2010.

[16] Wang. Y. Survey of Objective Video Quality Measurements.[on-line] Disponible en: <ftp://ftp.cs.wpi.edu/pub/techreports/pdf/06-02.pdf>. EMC Corporation Hopkinton, USA . 2004

[17] Botia.D, Gaviria. N, Jiménez. D, Menéndez. J.M. An approach to correlation of QoE metrics applied to VoD service on IPTV using a Diffserv Network.. 4th IEEE Latin-American Conference on Communications (IEEE LATINCOM 2012). Cuenca – Ecuador.2012.

[18] Botia.D, Gaviria. N, Botia. J, Jiménez. D, Menéndez. J.M. Improved the Quality of Experience Assessment with Quality Index Based Frames over IPTV Network. nd International Conference on Information and Communication Technologies and Applications ICTA 2012, Orlando – Florida. 2012.

[19] EvalVid, Evalvid: A Video Quality Evaluation Tool Set. [online] <http://www.tkn.tu-berlin.de/research/evalvid>. 2008

[20] Botia.D, Gaviria. N, J, Jiménez. D, Menéndez. J.M. Strategies for improving the QoE assessment over iTV platforms based on QoS metrics. el IEEE Workshop on Multimedia Computing Research. ISM2012. Irvine, California. ISBN: 978-1-4673-4370-1. DOI 10.1109/ISM.2012.98. 2012.

[21] QoE-RNN Tool. [on line] Disponible en <http://code.google.com/p/qoe-rnn/>. 2011

[22] DING.W, Tong.Y, Zhang. Q, Yang.D. "Image and video quality assessment using Neural Network and SVM". TSINGHUA SCIENCE AND TECHNOLOGY ISSN 1007-0214 18/19. Pp 112-116. Volumen 13, Number 1. Febrero 2008. China.

[23] Zhang.X, Wu.L, Fang.Y, Jiang. H. "A Study of FR Video Quality Assessment of Real Time Video Stream". (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No.6, 2012

[24] Kung,C-H, Yang.W-S, Huang.C-Y, Kung.C-M. "Investigation of the Image Quality Assessment using Neural Networks and Structure Similarity". Proceedings of the International Symposium on Computer Science, Vol. 2 Issue 1, p219. 2010

[25] Wang.C, Jiang.X, Meng.F, Wang.Y. "Quality Assessment for MPEG-2 Video Streams Using a Neural Network Model". IEEE 13th International Conference on Communication Technology (ICCT), pp 868 - 872. ISBN:978-1-61284-306-3. 2011

[26] Casas. P, Guerra. D, Bayarres.I . "Calidad de Servicio Percibida en Servicios de Voz y Video sobre IP". Facultad de Ingeniería. Universidad de la República. Proyecto de fin de carrera, Ingeniería Eléctrica- Plan 97, Telecomunicaciones. Agosto de 2005.

[27]Mohamed, S. Rubino, G. "A study of Real – Time Packet Video Quality Using Random Neural Networks". IEEE Trans. On Circuits and Systems for Video Technology. 12, pp 1071-1083, 2002.

[28] Rubino, G, Varela. M. Bonnin, J.M. Controlling Multimedia QoS in the Future Home Network Using the PSQA Metric. The Computer Journal . 2006.

[29] Video Quality Experts Group, "Final Report from the VQEG on the validation of Objective Models of Video Quality Assessment, Pase II", [on-line] Disponible en : <http://www.vpeg.org>, 2003.

[30] Winkler, S. Digital Video Quality – Vision, Models and Metrics. Ed Jhon Wiley & Sons. Suiza. 2005.

[31] Vatolin. D, "MSU Video Metric Quality Tool". MSU Graphics and media Lab [on-line]. Disponible en :http://compression.ru/video/quality_measure/video_measurement_tool_en.html. Rusia. 2002

[32] DSL Forum, "Triple-play Services Quality of Experience Requirements and Mechanism", Working text WT-126 version 0.5, Febrero 21, 2006.