

Incremento de la calidad de juegos en la nube utilizando técnicas de aprendizaje apilado

Mauricio Marín Author^{1,2*}, Jorge Gutierrez Author^{2,3†} and
Nicolás Cárdenas Author^{1,2†}

^{1*}Ing. Sistemas, Universidad Industrial de Santander, Puente Nacional,
684521, Santander, Colombia.

²Ing. Sistemas, Universidad Industrial de Santander, Monquirá, 154260,
Santander, Colombia.

³Ing. Sistemas, Universidad Industrial de Santander, Barbosa, 684511,
Santander, Colombia.

*Corresponding author(s). E-mail(s): marinmauricio1000@gmail.com;
Contributing authors: jorge1706guitierrez@gmail.com;
nicolascardenas311@gmail.com;

[†]These authors contributed equally to this work.

Abstract

Los juegos en la nube presentan ventajas rentables para los usuarios y desarrolladores al eliminar las inversiones en hardware y las compras de juegos, reducir los costos de desarrollo y optimizar los esfuerzos de distribución. Sin embargo, presentan desafíos para los operadores del servicio de Cloud Gaming, ya que exigen baja latencia y una potencia computacional sustancial. Por otro lado, la satisfacción del usuario en los juegos en la nube depende de varios factores, incluido el contenido del juego, el tipo de red y el contexto, todos los cuales configuran la calidad de la experiencia.

Keywords: nube, juegos, optimizar, redes

1 Introducción

En los últimos años, el rápido avance de la tecnología ha impulsado el crecimiento y la adopción de los servicios de juegos en la nube, proporcionando a los usuarios

la capacidad de jugar títulos de alta calidad sin la necesidad de hardware avanzado. Este paradigma transforma la experiencia del usuario, trasladando la carga de procesamiento desde los dispositivos locales hacia servidores remotos y altamente especializados.

La funcionalidad y el rendimiento de estos servidores en la nube son cruciales para asegurar una experiencia de juego fluida y satisfactoria. Este estudio se centra en la arquitectura de servidores y su influencia en el rendimiento de los servicios de juegos en la nube. Exploramos cómo los servidores optimizados con componentes de última generación, junto con modelos avanzados de inteligencia artificial, pueden mejorar la Calidad de la Experiencia (QoE) de los usuarios. Específicamente, se analiza el uso de servidores equipados con procesadores potentes, memoria de alta velocidad y tarjetas gráficas dedicadas para manejar las demandas intensivas de los juegos modernos. Además, se discute la implementación de algoritmos de aprendizaje automático para predecir y mejorar la QoE, considerando la variabilidad en la red y las preferencias del usuario.

A través de un análisis comparativo entre configuraciones de servidores y su impacto en la latencia, fluctuación y pérdida de paquetes, el estudio proporciona una visión integral de cómo la infraestructura soporta la creciente demanda de juegos en la nube.

Utilizamos Stacking Learning: Fue implementado con el fin de mejorar la generalización y la precisión del modelo. Este enfoque combina múltiples modelos de aprendizaje automático en un solo modelo, aprovechando las fortalezas de cada uno para obtener una mejor predicción general.

2 Metodología

El estudio comienza con un análisis de la distribución de los servidores de Cloud Gaming, tomando como referencia a GeForceNow de NVIDIA. Se realizará un resumen de cómo se optimizan los juegos en estos servidores para proporcionar una mejor experiencia a los usuarios. Esta optimización es crucial para garantizar una alta calidad de experiencia (QoE), un aspecto que se aborda en la segunda parte del estudio.

A continuación, se examinan diferentes modelos de aprendizaje para la medición de la QoE de los usuarios que participan en el Cloud Gaming (CG). Basándonos en trabajos previos como [1], se estudiará el impacto de estos métodos sobre las decisiones que se pueden tomar para mejorar o innovar en el Cloud Gaming.

2.1 Distribución de Servidores

Esta sección engloba el funcionamiento de la CPU y la adaptabilidad de los componentes dentro de la torre de una PC, destacando cómo estos elementos son cruciales para el rendimiento de los servidores utilizados. Se mencionará brevemente cómo está

organizado el hardware de los servidores para poder brindar a los usuarios experiencias de jugabilidad de alta calidad.

Empezando con los servidores que utiliza NVIDIA, los RTX Blade son un tipo de servidor más delgado y que consume menos energía. Cada tarjeta del servidor está compuesta por un procesador i9 que corre a 3.5 GHz de frecuencia, acompañado de 64 GB de memoria RAM, 1 TB de almacenamiento SSD M2 y dos tarjetas gráficas en paralelo. Todo esto se duplica, proporcionando una potencia y eficiencia excepcionales para soportar las demandas.

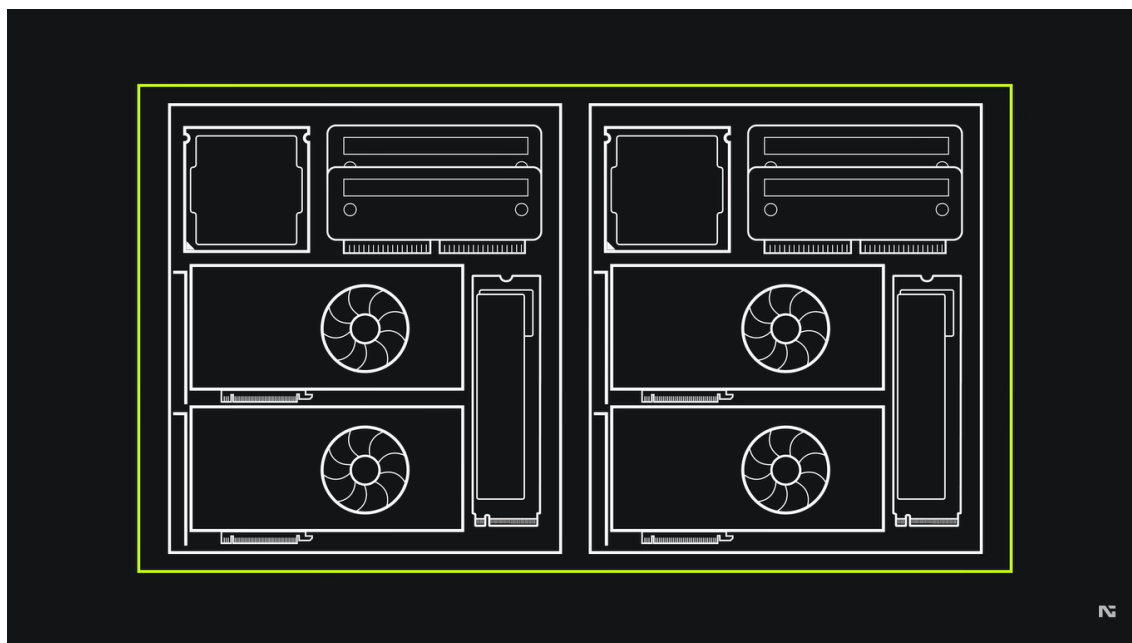


Fig. 1 Distribución de las RTX Blade de NVIDIA

Aunque este modelo no refleja con total precisión la realidad, se asemeja bastante. Para realizar la optimización de estas tarjetas, se crean instancias virtuales. Las instancias virtuales permiten una distribución eficiente de los recursos del servidor, asegurando que cada jugador reciba la potencia necesaria para su juego sin desperdiciar capacidad. Esta estrategia es crucial para mantener una alta calidad de experiencia (QoE), optimizando el uso del hardware y garantizando un rendimiento consistente para todos los usuarios.

Existen varios inconvenientes para generar una experiencia satisfactoria para los usuarios, como la transmisión de datos. Los usuarios desean jugar en tiempo real sin experimentar latencia alguna. Para lograrlo, se utilizan algoritmos de compresión

que, hoy en día, son ampliamente utilizados por plataformas como YouTube, Netflix, Amazon Prime, entre otras. Estos algoritmos eliminan información de cada fotograma enviado del juego, suprimiendo áreas que carecen de relevancia. Además, estos algoritmos analizan la calidad de la conexión y, si el ancho de banda es limitado, eliminan más datos de la imagen. Por eso, a veces notamos que la resolución de un video en YouTube disminuye cuando nuestra conexión es deficiente.



Fig. 2 Ejemplo de un Algoritmo de compresión

2.2 Modelos de referencia

Como se menciona en [2], se utilizaron varios modelos de deep learning denominados "Modelos base", los cuales se emplean para el preprocesamiento de los datos antes de que el meta modelo realice la estimación precisa de la QoE. Para esto, se utilizaron 3 conjuntos de datos: Wired, Mobile y Expanded.

A continuación, se describe de manera general cómo funciona el Stacking Learning y cómo los datos de salida de estos "Modelos base" se utilizan como entrada para el meta modelo encargado de predecir la QoE.

2.3 Stacking Learning

Este modelo consiste en un algoritmo de aprendizaje automático que puede mejorar el rendimiento del modelo de manera significativa. De manera resumida, el algoritmo combina varios modelos de aprendizaje automático para reducir la varianza, el sesgo o aumentar la precisión de la predicción [3].

Los autores dividen el modelo en 2 subniveles a los que denominan Nivel 0 y Nivel 1.

2.3.1 Nivel 0

Este subnivel consta de 2 o más modelos base, que pueden ser entrenados con una combinación de modelos no lineales como random forest, decision tree y XGBoost.

2.3.2 Nivel 1

Este subnivel no se describe explícitamente en el artículo, pero consiste en que las salidas generadas por los modelos base que utiliza el nivel 0 son las entradas para el siguiente modelo, denominado "meta-modelo". Este modelo se entrena con los datos ya procesados y genera la estimación para el QoE.

3 Resultados

3.1 Servidores

Como se explicó en la metodología, existen distintos retos para el CG, la experiencia del usuario puede verse afectada drásticamente por factores como:

- Latencia
- Resolución de imagen
- Conectividad

Estos son simplemente algunos de los factores a tener en cuenta. La estimación de QoE es una buena forma de observar el nivel de satisfacción de los usuarios, ya que todos estos factores están en juego al mismo tiempo. Algunos aspectos que no pueden controlarse se comunican directamente al usuario; por ejemplo, en las plataformas de Cloud Gaming siempre se les recomienda tener una velocidad de Internet (subida) de al menos 100 MB/s para evitar retrasos.

La estimación de QoE es importante para las decisiones empresariales, como invertir, seguir mejorando e incluso innovar en sus servicios para atraer más clientes y así generar una mejor rentabilidad. Algunas empresas podrían desarrollar sus propios estimadores para identificar áreas específicas de mejora y evaluar la satisfacción de los usuarios con la experiencia de juego.

3.2 Evaluación de Modelos y Comparación

Modelos Base y Stacking Learning: En el estudio realizado en [2], se utilizaron varios modelos de deep learning denominados "Modelos base" para el preprocesamiento de datos, y se implementó Stacking Learning para mejorar la precisión de las predicciones de QoE.

El enfoque de Stacking Learning mostró una mejora significativa en el Error Cuadrático Medio (MSE), reduciéndolo de 1.17 a 0.042. Esto indica una mayor precisión en las predicciones en comparación con los modelos individuales. La relación de precisión y el ratio de éxito se mantuvieron constantes, reflejando la consistencia del modelo expandido.

Generalización y Datos Descubiertos: El modelo de Stacking Learning expandido también se validó utilizando datos descartados, que incluían jugadores con participación limitada en los partidos. El rendimiento del modelo en estos datos refuerza su capacidad de generalización, proporcionando predicciones precisas incluso para casos menos comunes.

3.2.1 Importancia de las Características

El análisis de la importancia de las características utilizando técnicas como SHAP (Shapley Additive Explanations) reveló que algunas características, como el Video Jitter, son consistentemente importantes en la predicción de QoE. La importancia de estas características se mantuvo estable tanto en el modelo original como en el expandido, destacando su relevancia para la estimación precisa de la QoE.

3.2.2 Comparación con Modelos Individuales

Aunque los modelos individuales como XGBoost ofrecieron un rendimiento competitivo, el Stacking Learning demostró ser superior en términos de reducción del MSE. Esta ventaja subraya la efectividad de los enfoques de ensamblaje en el manejo de la complejidad y variabilidad de los datos de QoE.

La figura 3 muestra la distribución del QoE en los conjuntos de datos Wired, Mobile y Expanded. La distribución de QoE es más equilibrada al fusionarlos en el Expanded. Esto debería deberse al mayor número de coincidencias. En el conjunto de datos móviles, la mayoría de los jugadores evaluaron el QoE como extremadamente malo, a pesar de la distribución homogénea de los parámetros de la red, esto probablemente esté relacionado con la reacción inalámbrica a la fluctuación y la pérdida.

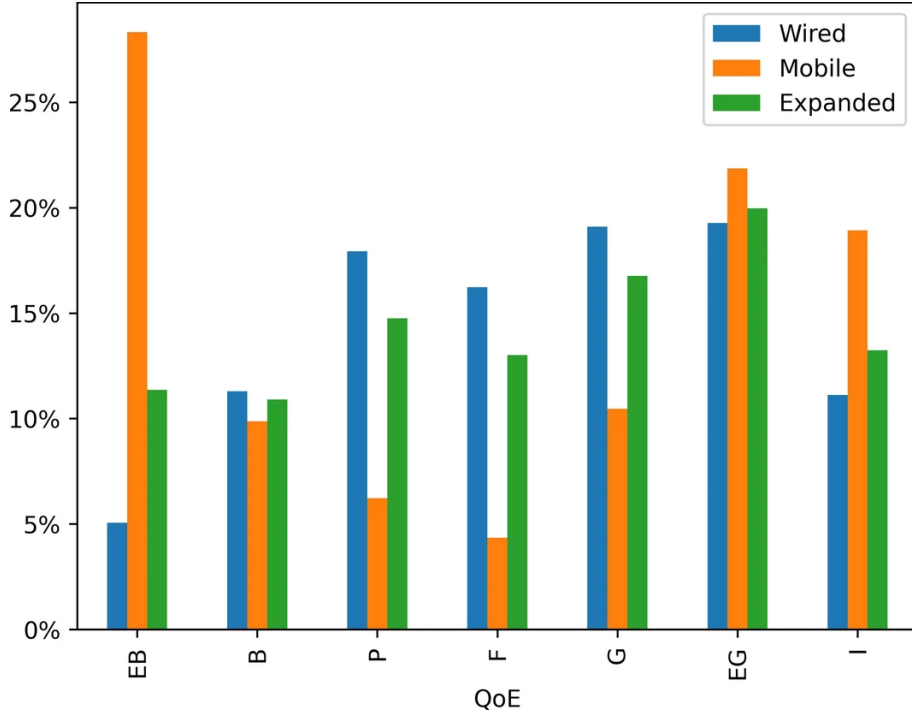


Fig. 3 Reprodutor QoE para cada conjunto de datos. EB extremadamente malo, B malo, P pobre, F justo, G bien, E excelente, I ideal

4 Conclusiones

Este estudio destaca la importancia de diversos factores críticos para el éxito de los servicios de Cloud Gaming (CG). La infraestructura de servidores juega un papel fundamental en la entrega de una experiencia de usuario satisfactoria, asegurando que millones de usuarios puedan conectarse simultáneamente sin interrupciones significativas. La implementación de instancias virtuales y balanceadores de carga es esencial para distribuir eficientemente los recursos del hardware, manteniendo una alta calidad de experiencia (QoE).

El análisis de la QoE es vital para comprender el nivel de satisfacción de los usuarios y orientar las decisiones empresariales. La aplicación de modelos de aprendizaje automático, como el Stacking Learning, ha demostrado mejorar significativamente la precisión de las predicciones de QoE. Este enfoque no solo reduce el Error Cuadrático Medio (MSE), sino que también ofrece una mayor capacidad de generalización, proporcionando predicciones precisas incluso en casos menos comunes.

Además, la distribución de QoE varía notablemente entre los diferentes contextos de red. En redes móviles, la QoE tiende a ser percibida como extremadamente mala, a pesar de una distribución homogénea de los parámetros de red. Esto resalta la

susceptibilidad de las conexiones inalámbricas a la fluctuación y pérdida de datos, afectando negativamente la experiencia del usuario.

References

- [1] Sabet, S.S., Schmidt, S., Zadtootaghaj, S., Griwodz, C., Möller, S.: Delay sensitivity classification of cloud gaming content, 25–30 (2020) <https://doi.org/10.1145/3386293.3397116>
- [2] Soares, D., Carvalho, M., Macedo, D.F.: Mejora de cloud gaming qoe estimación por stacking learning. J Netw Syst Administrar **32**, 58 (2024) <https://doi.org/10.1007/s10922-024-09836-6>
- [3] Elwerghemmi, R., Heni, M., Ksantini, R., Bouallegue, R.: Online qoe prediction model based on stacked multiclass incremental support vector machine, 1–5 (2019) <https://doi.org/10.1109/ICMSAO.2019.8880302>