

Agregando Información contextual para uso en sistemas de recomendación contextualizado: MusicSensory

1st Victor Quicaño Miranda

*Escuela Profesional de Ciencia de la Computación
Universidad Nacional de San Agustín
Arequipa, Perú
vquicano@unsa.edu.pe*

2nd Dra. Ana María Cuadros

*Escuela Profesional de Ciencia de la Computación
Universidad Nacional de San Agustín
Arequipa, Perú*

Abstract—MusicSensory surge de la combinación de tres fuentes de datos: Music4All, su versión Onion, y Extrasensory. Music4All proporciona registros de escucha de música y la metadata de las canciones, mientras que Extrasensory recopila datos de actividad y contexto diario de 60 usuarios a través de sus dispositivos.

La idea fue simular que estos usuarios estaban escuchando música mientras realizaban sus actividades diarias. Para lograrlo, se sincronizaron los datos usando las marcas de tiempo (timestamps), primero emparejando los días de la semana y luego coincidiendo la hora y el minuto del día.

Index Terms—Recommender Systems, Data Visualization, Context-awareness

I. INTRODUCCIÓN

Los **Sistemas de recomendación** son sistemas inteligentes o mecanismos de software cuya principal función es ofrecer sugerencias sobre ítems o diferentes tipos de entidades, como por ejemplo: Productos, servicios o contactos [1], [2]. El objetivo es dar a conocer aquellos ítems que podrían ser útiles, interesantes o más relevantes para un usuario en específico. En esencia, un sistema de recomendación presenta información a los usuarios sobre una serie de ítems [3], que basándose en sus preferencias, se adaptan, son predichos, siendo así los más apropiado para ellos. La importancia de esta tecnología recae en que aliviana la carga de información que el usuario recibe [4], permitiéndoles encontrar o descubrir fácilmente aquellos ítems adecuados a sus gustos. Los propósitos de un sistema de recomendación, por ejemplo en ámbito comercial, pueden incluir aumentar el número de ítems vendidos, vender ítems más diversos y aumentar la satisfacción y fidelidad del usuario [2].

El funcionamiento de un sistemas de recomendación es simple, en una primera instancia se recopilan información de las preferencias de los usuarios para un conjunto de ítems [2]. Esta información se puede obtener de forma explícita, como registrando las calificaciones de los usuarios, o implícitamente, a través del comportamiento del usuario (Tiempo de visualización, número de búsquedas, búsquedas similares a dicho

ítem, etc) [4], [3]. Con base en estos datos, los sistemas aprenden las selecciones de los usuarios para establecer patrones de relación. La parte central de un sistema de recomendación suele ser la regresión de calificaciones, que es una predicción de la calificación que el sistema cree que el usuario dará. Predicen la calificación de los ítems a recomendar al usuario y crean una lista o ranking de recomendaciones para cada usuario [5].

Tradicionalmente, los sistemas de recomendación tratan aplicaciones con solían basarse solo en dos tipos de entidades: usuarios e ítems, sin considerar el contexto al proporcionar recomendaciones [6], [1]. Su entrada típica se basa en registros de la forma {usuario, ítem, calificación}. Sin embargo, actualmente un avance significativo ha sido la incorporación del contexto como un conjunto de atributos adicionales para mejorar la precisión de las preferencias del usuario. Los CARSS suelen manejar registros de datos de la forma {usuario, ítem, contexto, calificación} [6].

Teniendo en cuenta lo anterior también es posible clasificar los sistemas de recomendación según la técnica de estimación de calificaciones que utilizan:

- Por filtrado colaborativo (CF): Trabajan recopilando las opiniones de los usuarios en forma de calificaciones y explotan las similitudes en el comportamiento de calificación entre varios usuarios para determinar cómo recomendar un ítem.
- Basados en contenido: Recomiendan ítems con atributos similares a los que le gustan al usuario, basándose en la información de los ítems.

También existen sistemas híbridos que combinan diferentes enfoques [1], [7].

Siendo así una tecnología fundamental en sitios web de comercio electrónico, y se han expandido y utilizado en diversos campos de servicio como streaming, redes sociales, turismo, atención médica (healthcare), educación y información académica. Algunos ejemplos de aplicaciones prácticas son la recomendación de libros, CDs en Amazon.com, películas por Movielens, música por last.fm y noticias en VERSIFI technologies [7]. Siendo la música uno de los contenidos

de estudio abordados por los sistemas de recomendación, ya que con su capacidad para evocar emociones profundas y universales, ha sido durante siglos un medio esencial de expresión humana. Las canciones, en particular, combinan letra y melodía para crear una experiencia emocional única, capaz de influir en el estado de ánimo y los sentimientos de los oyentes [8].

El problema que aborda esta investigación se da apartir de limitaciones en los sistemas actuales de recomendación musical, en los predominan enfoques que se basan en el historial de reproducción del usuario o en métricas de similitud entre elementos musicales. La forma tradicional de hacer esto no toma en cuenta suficientes tipos diferentes de datos. No considera cosas importantes como el contexto, las emociones o la situación del momento, que influyen mucho en la música que le gusta a una persona.

Cuando los sistemas no consideran estos factores, las recomendaciones musicales pueden parecer fuera de lugar. A veces resultan repetitivas, poco relevantes o incluso molestas, porque no se ajustan a lo que la persona necesita emocionalmente o a la situación que está viviendo. Al no incluir las emociones y los sentimientos como parte del análisis, se pierde la oportunidad de ofrecer una experiencia realmente personalizada y significativa.

Por ello, el objetivo principal en este trabajo consiste en desarrollar un dataset musical contextualizado que integre de manera multimodal variables contextuales, ubicación, emoción endebida dentro de la letra y el tiempo de escucha del usuario, abordando la heterogeneidad de datos necesaria para mejorar la calidad y relevancia de los sistemas de recomendación musical mediante la incorporación sistemática del contexto situacional, las emociones y los sentimientos del usuario.

II. TRABAJOS RELACIONADOS

A. Con Respecto al modelo

1) *Aggregating Contextual Information for Multi-Criteria Online Music Recommendations* [9]: El artículo presenta CAMCMusic un aporte significativo en el ámbito de los sistemas de recomendación musical, particularmente cuando se analiza desde la perspectiva del aprendizaje federado y la personalización contextual. En un entorno donde las preferencias musicales están profundamente influenciadas por factores subjetivos y variables contextuales, CAMCMusic ofrece un enfoque innovador al integrar la toma de decisiones multicriterio (MCDM) con una conciencia contextual dinámica. Esta propuesta representa un paso importante hacia la superación de las limitaciones de los sistemas tradicionales que, por lo general, dependen de calificaciones explícitas, atributos fijos del usuario o características estáticas de las canciones. La capacidad de CAMCMusic para generar recomendaciones relevantes sin necesidad de información directa del usuario abre nuevas posibilidades para desarrollar sistemas de recomendación más privados y escalables.

Desde la óptica del aprendizaje federado, CAMCMusic se alinea con los principios de minimizar la necesidad de recopilar información sensible directamente del usuario. Aunque el

artículo no implementa explícitamente un enfoque federado, su estrategia de utilizar datos preexistentes y recomendaciones estereotipadas se puede complementar eficazmente con técnicas de aprendizaje federado para proteger la privacidad del usuario. Al combinar estos dos enfoques, se podría fortalecer la personalización musical respetando los principios de privacidad diferencial y aprendizaje descentralizado. Este vínculo es especialmente relevante en un contexto donde la recopilación de datos emocionales, de ubicación o estado físico puede presentar riesgos éticos y legales si no se maneja cuidadosamente.

Además, la relevancia de CAMCMusic se magnifica al abordar el problema del "arranque en frío", uno de los desafíos persistentes en los sistemas de recomendación. Al evitar depender exclusivamente de información previa del usuario, y al construir recomendaciones basadas en patrones de comportamiento contextual agrupados, el sistema no solo mejora la calidad de las recomendaciones, sino que también facilita la adopción en escenarios con datos limitados o usuarios nuevos. Esta capacidad puede aprovecharse dentro de entornos de aprendizaje federado para hacer recomendaciones útiles sin compartir datos personales entre dispositivos o servidores.

Finalmente, CAMCMusic sienta las bases para el desarrollo de futuros sistemas de recomendación que integren el aprendizaje federado con una estructura multicriterio contextual. Esta combinación permitiría una evolución significativa en la personalización musical: sistemas más seguros, adaptativos y éticamente sostenibles que comprenden mejor las complejidades del comportamiento humano sin comprometer la privacidad. Por lo tanto, el artículo no solo contribuye al estado del arte en recomendación musical, sino que también abre caminos hacia una integración más profunda entre inteligencia artificial, privacidad y experiencia del usuario.

B. Con respecto al tipo de datos

1) *Listener Modeling and Context-Aware Music Recommendation Based on Country Archetypes* [10]: Este artículo representa una contribución destacada al campo de los sistemas de recomendación musical al incorporar de manera efectiva el contexto geográfico del usuario en modelos avanzados de aprendizaje profundo. Su enfoque innovador parte del análisis de grandes volúmenes de datos de comportamiento musical y propone una arquitectura de autoencoder variacional (VAE) mejorada que integra información contextual sin depender de fuentes externas. Esta estrategia no solo optimiza la personalización de las recomendaciones, sino que también mantiene un enfoque auto-sostenible, lo cual resulta especialmente valioso para plataformas que buscan soluciones escalables y eficientes sin requerir datos sensibles o difíciles de obtener.

Desde la perspectiva del aprendizaje federado, la propuesta del artículo abre caminos para el desarrollo de sistemas descentralizados y respetuosos con la privacidad. Aunque el modelo actual no aplica directamente técnicas federadas, su dependencia exclusiva del historial de escucha y del país autoinformado por el usuario facilita una posible implementación

futura en entornos federados. En tales escenarios, los modelos podrían entrenarse localmente en los dispositivos de los usuarios, manteniendo su información personal segura mientras se sigue aprovechando el contexto geográfico para mejorar la calidad de las recomendaciones. Esto es especialmente relevante dado el creciente énfasis en la privacidad de los datos en la inteligencia artificial moderna.

Además, la investigación subraya un aspecto que a menudo es pasado por alto: la importancia del contexto cultural y geográfico en las preferencias musicales. A través de la identificación de arquetipos de países mediante técnicas no supervisadas, los autores logran capturar patrones colectivos de escucha sin imponer supuestos culturales explícitos. Esta abstracción de alto nivel es particularmente útil para construir modelos más generalizables que pueden adaptarse a distintos grupos de usuarios sin requerir personalización individual exhaustiva. La integración de estos arquetipos en una arquitectura VAE con mecanismos de *gating* permite que las preferencias individuales sean moduladas por el entorno colectivo, reflejando una forma sofisticada de conciencia contextual.

En conjunto, el artículo no solo ofrece un avance técnico en términos de modelado de usuario consciente del contexto, sino que también propone una estructura sólida que puede evolucionar hacia entornos de aprendizaje federado. En el contexto más amplio del desarrollo de sistemas de recomendación éticos, escalables y personalizados, esta investigación proporciona herramientas y perspectivas clave para diseñar tecnologías que equilibran efectividad, adaptabilidad y privacidad. La combinación de geolocalización auto-informada, análisis de comportamiento y aprendizaje profundo representa una fórmula poderosa para futuras soluciones en la recomendación musical moderna.

III. VISIÓN GENERAL

A. Descripción de Dataset

1) *Music4All - Onion* [11]: El dataset Music4All-Onion es una extensión multimodal y a gran escala del dataset Music4All. Este fue diseñado para suplir las limitaciones comunes en los sistemas de recomendación musical actuales: la dependencia exclusiva de interacciones usuario-item (colaborative filtering), ignorando el contenido musical como base para la afinidad del usuario. Esta omisión se debe principalmente a la falta de datasets estandarizados que integren tanto información colaborativa como de contenido.

Para solucionar este problema, Music4All-Onion incorpora 26 nuevas características de tipo audio, video y metadatos, además de proporcionar 252,984,396 registros de escucha pertenecientes a 119,140 usuarios, extraídos de Last.fm. Esto permite la investigación comparativa del impacto de distintas capas semánticas del contenido musical (modelo tipo "cebolla") sobre la precisión, novedad y equidad de los sistemas de recomendación.

B. Music4All [12]

1) *Contexto*: El dataset Music4All surge como una contribución a la comunidad de Music Information Retrieval

(MIR), cuyo objetivo principal es el desarrollo de sistemas que permitan recuperar y recomendar contenido musical de manera eficiente y efectiva. Pese a los avances en el área, existía una carencia de bases de datos que cumplieran con ciertos requisitos fundamentales para la investigación:

- Disponibilidad de señales de audio
- Diversidad de atributos musicales
- Un volumen considerable de piezas musicales.

Es así que surge el DataSet Music4All, desarrollado para cubrir dichas necesidades. Dentro de sus bondades ofrece un conjunto de datos amplio y diverso que incluye información de más de **109,000 canciones** y el historial de escucha de **15,602 usuarios anónimos**. El dataset contiene diversos tipos de información para cada canción, incluyendo metadatos, clips de audio de 30 segundos, letras, información de popularidad y atributos acústicos derivados de la API de Spotify, así como etiquetas de género y tags generados por usuarios en Last.fm.

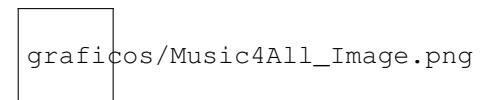


Fig. 1. Pipeline de la elaboración de Music4All.

2) *Entidad de estudio*: La entidad principal de estudio en Music4All es la canción. Cada canción está descrita por una serie de atributos que pueden agruparse en categorías como metadatos, características acústicas, y etiquetas semánticas **[Revisar la estructura del dataset en el Cuadro: ??]**

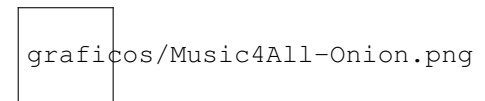


Fig. 2. Capas de Music4All-Onion.

3) *Entidad de estudio*: Music4All-Onion organiza la información del contenido musical en capas (layers), que representan distintos tipos de características semánticas. La entidad de estudio principal continúa siendo la canción, pero se analiza desde múltiples perspectivas (audio, texto, imagen, interacción):

- 1) **Capa de Audio (Audio Layer)** Contiene descriptores extraídos de la señal de audio con herramientas como OpenSMILE y Essentia. **[Revisar el Cuadro: ??]**
- 2) Capa de Contenido Derivado - Texto (Extracted Metadata/Derived, EMD) **[Revisar el Cuadro: ??]**
- 3) Capa de Contenido Generado por el Usuario (User-Generated Content, UGC) **[Revisar el Cuadro: ??]**
- 4) Capa de Contenido Derivado Visual (Derived Content, DC) **[Revisar el Cuadro: ??]**
- 5) Eventos de Escucha (Interacciones Usuario-Canción) **[Revisar el Cuadro: ??]**

C. Music4All [12]

El dataset Music4All surge como una contribución a la comunidad de Music Information Retrieval (MIR), cuyo objetivo principal es el desarrollo de sistemas que permitan recuperar y recomendar contenido musical de manera eficiente y efectiva. Pese a los avances en el área, existía una carencia de bases de datos que cumplieran con ciertos requisitos fundamentales para la investigación:

- Disponibilidad de señales de audio
- Diversidad de atributos musicales
- Un volumen considerable de piezas musicales.

Es así que surge el DataSet Music4All, desarrollado para cubrir dichas necesidades. Dentro de sus bondades ofrece un conjunto de datos amplio y diverso que incluye información de más de **109,000 canciones** y el historial de escucha de **15,602 usuarios anónimos**. El dataset contiene diversos tipos de información para cada canción, incluyendo metadatos, clips de audio de 30 segundos, letras, información de popularidad y atributos acústicos derivados de la API de Spotify, así como etiquetas de género y tags generados por usuarios en Last.fm.

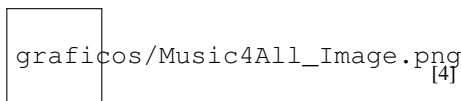


Fig. 3. Pipeline de la elaboración de Music4All.

Music4All nos brinda información simple y sencilla sobre las canciones que posee. Cada canción está descrita por una serie de atributos que pueden agruparse en categorías como metadatos, características acústicas, y etiquetas semánticas [Revisar la estructura del dataset en el Cuadro: ??]

1) *Extrasensory* [13]: El dataset ExtraSensory fue recopilado “en la vida real” mediante la aplicación móvil ExtraSensoryApp desarrollada en la UC San Diego. Durante aproximadamente un año (2015–2016), 60 voluntarios llevaron su smartphone y en algunos casos, también un smartwatch configurados para registrar automáticamente muestras de datos sensoriales cada minuto. Al mismo tiempo, los usuarios auto-reportaban su actividad y su contexto (por ejemplo, “caminando”, “sentado”, “en reunión”, “en casa”) mediante una interfaz de la app. Este diseño permite estudiar el reconocimiento de contexto humano y la actividad diaria bajo condiciones naturales, superando las limitaciones de los experimentos controlados en laboratorio

Este dataset dispone de **muestras de contexto**, dentro de un vector de características sensoriales y de estado asociado a un instante (1 min) y al usuario que lo generó. Cada muestra incluye: [Revisar el Cuadro ?? para conocer la estructura del Dataset]

- Metadatos de identificación
 - uuid – Identificador único de usuario (60 IDs)
 - timestamp – Marca temporal de la muestra

- Etiquetas de contexto
 - 50 columnas binarias, una por cada etiqueta (actividad o situación), con valor 1 si aplica, 0 si no.
 - Ejemplos: label_walking, label_sitting, label_meeting, label_home, etc
- Características sensoriales y de estado
 - Un total de 225 features agrupadas en seis grandes modalidades: acelerómetro, giroscopio, magnetómetro, audio, estado del teléfono y brújula del smartwatch.
 - Para cada sensor se extraen estadísticas temporales (media, mediana, desviación estándar, percentiles,...)

REFERENCES

- [1] L. Sharma and A. Gera, “Ijett - a survey of recommendation system: Research challenges,” *International Journal of Engineering Trends and Technology - IJETT*, 2013.
- [2] S. M. M. Seyednezhad, K. N. Cozart, J. A. Bowllan, and A. O. Smith, “A review on recommendation systems: Context-aware to social-based,” Nov. 2018. [Online]. Available: <https://arxiv.org/abs/1811.11866>
- [3] P. Mateos and A. Bellogín, “A systematic literature review of recent advances on context-aware recommender systems,” *Artificial Intelligence Review*, vol. 58, no. 1, Nov. 2024. [Online]. Available: <http://dx.doi.org/10.1007/s10462-024-10939-4>
- [4] H. Ko, S. Lee, Y. Park, and A. Choi, “A survey of recommendation systems: Recommendation models, techniques, and application fields,” *Electronics (Basel)*, vol. 11, no. 1, p. 141, Jan. 2022.
- [5] A. Gutiérrez Fandiño, “Explainability on recommender systems using contextual data,” Master’s thesis, Universitat Politècnica de Catalunya, 2020.
- [6] G. Adomavicius, B. Mobasher, F. Ricci, and A. Tuzhilin, “Context-aware recommender systems,” *AI Mag.*, vol. 32, no. 3, pp. 67–80, Sep. 2011.
- [7] H. Ko, S. Lee, Y. Park, and A. Choi, “A survey of recommendation systems: Recommendation models, techniques, and application fields,” *Electronics*, vol. 11, no. 1, p. 141, Jan. 2022. [Online]. Available: <http://dx.doi.org/10.3390/electronics11010141>
- [8] A. F. Alaminos-Fernández, “La música como lenguaje de las emociones. un análisis empírico de su capacidad performativa,” 2014.
- [9] J. Liu, “Aggregating contextual information for multi-criteria online music recommendations,” *IEEE Access*, vol. 13, pp. 8790–8805, 2025.
- [10] M. Schedl, C. Bauer, W. Reisinger, D. Kowald, and E. Lex, “Listener modeling and context-aware music recommendation based on country archetypes,” *Front. Artif. Intell.*, vol. 3, p. 508725, 2020.
- [11] M. Moscati, E. Parada-Cabaleiro, Y. Deldjoo, E. Zangerle, and M. Schedl, “Music4all-onion – a large-scale multi-faceted content-centric music recommendation dataset,” in *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, ser. CIKM ’22. ACM, Oct. 2022, p. 4339–4343. [Online]. Available: <http://dx.doi.org/10.1145/3511808.3557656>
- [12] I. A. Pegoraro Santana, F. Pinhelli, J. Donini, L. Catharin, R. B. Mangolin, Y. M. e. G. da Costa, V. Delisandra Feltrim, and M. A. Domingues, “Music4all: A new music database and its applications,” in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, Jul. 2020, p. 399–404. [Online]. Available: <http://dx.doi.org/10.1109/IWSSIP48289.2020.9145170>
- [13] Y. Vaizman, K. Ellis, G. Lanczkiet, and N. Weibel, “ExtraSensory app,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, Apr. 2018, pp. 1–12.