



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
ESCUELA DE INGENIERÍA  
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN

IIC3633 — Sistemas Recomendadores — 2' 2020

# Propuesta Proyecto

## 1 Contexto

Los sistemas de recomendación tienen como objetivo presentarle a un usuario una recomendación de ítems con el fin de ayudarlo a seleccionar aquellos de su interés. Puede ayudarlo a elegir qué canción escuchar, qué video ver, qué producto comprar, entre otros. Generalmente el subconjunto de ítems que el sistema le recomienda al usuario se extrae de un conjunto de gran tamaño que puede ser difícil de explorar. Ejemplos de estos grandes conjuntos de ítems son los catálogos música de Spotify, videos de YouTube y la cartelera de Netflix. Es por este motivo que los sistemas de recomendación son una herramienta que permite al usuario encontrar fácilmente el contenido que quiere consumir.

Sin embargo, los avances en inteligencia artificial, especialmente en redes neuronales generativas, han abierto una nueva posibilidad: generar nuevos ítems como recomendación para el usuario. Este método utiliza el potencial de las redes neuronales generativas como las GAN para realizar recomendación basada en contenido. Este proyecto busca explorar este horizonte mediante la implementación de un modelo generativo capaz de generar nuevos ítems de interés para el usuario.

## 2 Problema y su justificación

### 2.1 Problema

El problema que abordará el proyecto es la generación de música como recomendación basada en contexto. El contexto puede corresponder a una imagen, video, escenario de videojuego u otras piezas musicales. El objetivo es recomendar música generada que tome en consideración tal contexto y genere ideas musicales acorde a este.

### 2.2 Justificación

El problema seleccionado es relevante en diversas aplicaciones en las que se puede recomendar una pieza musical nueva. Se puede utilizar como audio libre de copyright para acompañar video o imágenes, como banda sonora para videojuegos dado el escenario, como herramienta de composición al sugerir alternativas dada ideas musicales previas o simplemente para ofrecer música completamente nueva a un usuario dada música de su interés.

### 2.3 Estado del arte

En cuanto a generación de música, existe literatura acerca de generación de música en basada en algún input como lyrics [2], emociones [5], expresiones faciales [3], o estilos [10]. Si bien estas investigaciones tienen casos de uso distintos a lo que se busca con el proyecto, sirven de aproximación para generar música basándose en

elementos externos a la música.

Por otro lado, existen múltiples papers de recomendación basada en contexto [6] e incluso para recomendación de música en específico basado en el contexto emocional del usuario [7]. De esta manera, existe ya un trabajo previo en recomendación basada en contexto, pero utilizan el enfoque más tradicional de recomendar música ya existente, mientras que la propuesta de este proyecto busca generar la música que va a ser recomendada.

No hay mucha investigación en cuanto a la generación de contenido para recomendación, pero existe un proyecto de una versión pasada del curso que ataca este problema con un enfoque en generación de arte para usuarios [11].

### 3 Objetivos

Nuestro objetivo en este proyecto es crear un modelo capaz de generar música acorde a un determinado contexto, inspirandonos en el enfoque contextual de los sistemas recomendadores. Nos interesa que la música generada efectivamente tome en cuenta la información contextual al momento de ser generada, con el fin de que sea una buena recomendación en perspectiva de su grado de relación con el contexto y secundariamente, su carácter musical.

La música a generar tiene como objetivo principal que admita información contextual en su generación y que efectivamente sea utilizada para ajustarla a tal contexto. En cuanto al carácter musical se exigirá que tenga coherencia rítmica y melódica de modo que un usuario pueda percibir las secuencias generadas como una pieza musical.

### 4 Solución propuesta

Nuestra solución preliminar para el problema consiste en la implementación de una red neuronal como LSTM que sea capaz de generar música que tenga la capacidad de recibir como input información acerca del contexto para el cuál se esta generando dicha música. La forma de incluir dicha información de momento esta por definir, pero puede ser de manera similar a como lo hacen los papers mencionados en la sección de estado del arte con las lyrics o las expresiones faciales.

Por otro lado, una parte importante del trabajo corresponderá a obtener dicha información contextual a partir de, por ejemplo, una imagen. Para esto tenemos múltiples ideas, que van desde simplemente codificar dicha imagen a un vector, como también quizás extraer las features de estilo como se hace en trabajos de style transfer [8] y que el contexto en este caso sea el estilo artístico de la imagen para la cual se quiere generar música.

### 5 Descripción de experimentos a realizar

#### 5.1 Experimentación

Dado a que el proyecto tiene un enfoque investigativo, hay varias decisiones sin tomar que requieren de experimentación:

1. Arquitectura del generador: En primer lugar se tiene que probar distintas arquitecturas y modelos para el generador de música, con el fin de seleccionar la más adecuada para esta tarea. Un buen punto de partida es considerar modelos como LSTM [10], la versión LSTM-GAN y los modelos generativos de Google Magenta [9]. A partir de estos, se tiene que modificar su estructura para incluir un input para el contexto y otras posibles cambios para que el modelo cumpla el objetivo propuesto.

2. Formato del contexto: Actualmente, el concepto de contexto que se esta empleando es abstracto. Se requiere de experimentación para orientar de qué manera vamos a representar el contexto de la imagen. Posibles alternativas son: análisis de sentimiento de la imagen, embeddings de la imagen completa, o features de estilo.
3. Datos de entrenamiento: Se debe decidir qué datos de música se le proporcionarán al modelo para su entrenamiento. Preliminarmente se planea entregarle música en formato MIDI al generador para entrenarlo, pero queda por definir que pistas específicamente le vamos a entregar (por ejemplo entregar ejemplos de determinados géneros musicales). También se podría considerar entregar más datos para complementar los archivos MIDI como metadata de canciones.

## 5.2 Métricas de evaluación

Dado que el objetivo comprende generar nueva música en vez de recomendar otras ya existentes, las métricas de recomendación tradicional no serán de utilidad. En cambio, hay dos aspectos para los cuales se deben implementar métricas con el fin de asegurar que nuestro objetivo se este cumpliendo:

1. Musicalidad: Las métricas relacionadas a musicalidad deben medir que tan musical son las secuencias generadas. Este punto es muy importante ya que es de interés la generación de música, y sin esto podría ocurrir que se genere un output que dadas las métricas de contexto sea muy bueno, pero que simplemente no sea interpretado como música por el usuario.
2. Adecuación al contexto: Las métricas relacionadas a la adecuación al contexto medirían que tan adecuada es la música generada para el contexto que se dió como input. Esta métrica es uno de los puntos clave de la investigación, pues será las que nos permitirá medir si efectivamente estamos creando música condicionada a una situación (en este caso imagen) o no.

## 5.3 Datasets

- Musical AI MIDI Dataset: <https://composing.ai/dataset>
- The Lakh MIDI Dataset v0.1: <https://colinraffel.com/projects/lmd/>
- FMA: A Dataset For Music Analysis: <https://github.com/mdeff/fma>
- Painter by Numbers: <https://www.kaggle.com/c/painter-by-numbers/data>

## 6 Otras ideas de proyecto

Dado que se permitió poner más ideas a parte de la principal para el proyecto, mencionaremos brevemente aquí otras ideas que tenemos para el proyecto para así poder tener feedback de cuál podría ser mejor considerando que nuestra intención es investigar un problema que se pueda publicar en un paper:

- Continuar la investigación en generación de arte para recomendación [11], enfocándonos en mejorar la calidad del arte generado. Tal como describen en el paper, el modelo que crearon se puede mejorar dado que las imágenes que se generan todavía son de baja calidad, para lo cual sería útil explorar arquitecturas más modernas de GANs y entrenar con un dataset más grande. Además, nos gustaría utilizar nociones de estilo artístico al momento de generar (Tal como se hace en la StyleGAN [12]). De esta manera, podríamos saber que pinturas se utilizaron como referencia artística para la creación de una determinada obra de arte.
- Recomendación de música para acompañar imágenes basada en el estilo artístico de la imagen: Esta idea consiste en recomendar música que vaya acorde al estilo de una imagen, utilizando técnicas de extracción de estilo para obtener una representación de la imagen, y luego con ese input recomendar

música que la acompañe de buena manera. A diferencia de nuestro proyecto actual, aquí no estaríamos generando música sino que solo recomendando canciones ya existentes, por lo que sería de menor complejidad que el proyecto propuesto. Aún así es un problema difícil, ya que no hay una métrica absoluta para decidir si una canción va bien con el estilo artístico de una imagen, ya que de por sí son dominios muy distintos.

- Generación de música para recomendación dado texto en vez de imágenes como contexto. Esto sería como para acompañar la lectura, de manera que pueda generar música que vaya acorde con el contenido que se este leyendo. Este acercamiento de generación de música puede ser más factible de realizar que el planteado en la propuesta principal de proyecto, dado que las palabras entregan información mucho más precisa acerca del contexto y del sentimiento del texto, en especial con modelos como BERT que son excelentes para el procesamiento del lenguaje natural. De esta manera, evaluar si una música va acorde al contexto o no va a ser más sencillo considerando que ya existe trabajo previo que genera música en base a emociones (Aunque siempre se ha hecho con emociones en caras antes que en texto). Además, las aplicaciones de esto podrían ser muy interesantes, como generación en línea de música mientras uno lee un libro digital, con el fin de mejorar la experiencia de lectura.

Por otro lado se podría directamente quitar la componente de generación y simplemente recomendar música basado en un texto y quizás en gustos del usuario con el fin de simplificar un poco el proyecto, pero no sabemos si eso tendría suficiente contribución para ser publicable. Esta aplicación también es interesante, pues recomendaría en base a los gustos de un usuario pero también en base al contexto en que se encuentra, que en este caso sería la lectura de un libro con un determinado contenido, el cuál se toma en cuenta al momento de generar la recomendación.

- Enfocar la recomendación de música generada para una aplicación en la composición. Dada secuencias previamente construidas por el usuario, diseñar un sistema que le ofrezca alternativas para continuar su creación ofreciendo un abanico de posibilidades desde propuestas más conservadoras a otras más arriesgadas. Se puede aplicar en melodías o ritmos mediante secuencias MIDI o armonías con progresiones de acordes [1].

Solicitamos respetuosamente feedback para estas ideas dado que nos es de interés hacer un proyecto publicable. Estamos conscientes que el proyecto presentado en la propuesta puede ser muy difícil de realizar en este tiempo, por lo cuál nos serviría saber qué tan buenas son las otras ideas presentadas. Muchas gracias de antemano.

## 7 Bibliografía relevante

### References

- [1] Huang, Cheng-Zhi Anna. 2019. *Deep Learning for Music Composition: Generation, Recommendation and Control*. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences. Michel Goossens, Frank Mittelbach, and Alexander Samarin.
- [2] Yi Yu, Simon Canales. 2019. *Conditional LSTM-GAN for Melody Generation from Lyrics*.
- [3] Pooja Mishra, Himanshu Talele, Yogesh Sawarkar, Rohit Vidhate, Ganesh Naikare. 2019 *Music Tune Generation based on Facial Emotion*. International Journal of Engineering Research & Technology (IJERT)
- [4] Gwenaelle C. Sergio, Rammohan Mallipeddi, Jun-Su Kang, Minhoo Lee. 2015. *Generating Music from an Image* Kyungpook National University, Daegu, South Korea.

- [5] Madhok, R., Goel, S. and Garg, S. *SentiMozart: Music Generation based on Emotion*. In Proceedings of the 10th International Conference on Agents and Artificial Intelligence (ICAART 2018)
- [6] Pazzani, M. J., Billsus, D. (2007). *Content-based recommendation systems*. In The adaptive web (pp. 325-341). Springer Berlin Heidelberg. Xu, W., Liu, X., Gong, Y. (2003).
- [7] Chih-Ming Chen, Ming-Feng Tsai, Jen-Yu Liu, Yi-Hsuan Yang. 2013. *Using Emotional Context from Article for Contextual Music Recommendation*
- [8] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge. 2015. *A Neural Algorithm of Artistic Style*.
- [9] Adam Roberts, Jesse Engel, Colin Raffel, Curtis Hawthorne, Douglas Eck. 2019. *A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music* Proceedings of the 35th International Conference on Machine Learning.
- [10] Huanru Henry Mao, Taylor Shin, Garrison W. Cottrell. 2018. *DeepJ: Style-Specific Music Generation*
- [11] Eugenio Herrera and Tamara Cucumides. 2020. *Personalized Artwork Generation for Recommendation*. In Proceedings of ACM Conference (Conference' 17).ACM, NewYork,NY, USA, 6 pages. <https://doi.org/10.1145/nnnnnnnn. nnnnnnnn>
- [12] Karras, Tero and Laine, Samuli and Aila, Timo. 2019 *A Style-Based Generator Architecture for Generative Adversarial Networks* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)