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

Dentro del contexto mencionado anteriormente, nos interesa atacar el problema de generación de música con el fin de poder recomendarla. Dicha música estaría generada tomando en cuenta un determinado contexto, que puede ser un usuario con sus gustos de música, un videojuego, un video, una imagen, entre otros. La idea es que la música generada vaya acorde al contexto seleccionado, por ejemplo para el caso de un usuario se debe generar música que le guste, para el videojuego se debe generar música que vaya bien con la situación del juego como también con el resto de banda sonora que pueda o no haber, y así con el resto.

2.2 Justificación

El problema mencionado anteriormente puede ser relevante para muchas aplicaciones en la vida real, que van desde simplemente recomendar música nueva que no exista antes a una persona para que pueda escuchar, como también generar audio libre de copyright para acompañar a un video, o incluso generar pistas de banda sonora para escenarios de videojuegos o películas. Además de lo anterior, también podría servir como inspiración para compositores en el momento de creación músical, recomendandole pistas acordes a un contexto para así dar pie a nuevas ideas de composición.

2.3 Estado del arte

En cuanto a generación de música, existen papers que generan música en base a cierto input como pueden ser lyrics [2], emociones [5], en expresiones faciales [3], o estilos [10]. Si bien esto se acerca a nuestra idea, apuntan a un caso de uso distinto al que nosotros queremos atacar, el cuál es un determinado contexto. Aún así, los métodos utilizados para generar música influenciada por ciertos factores nos puede servir.

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 nuestra propuesta 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.
La música a la que apuntamos es más de "ambiente", y que sea relativamente sencilla, como por ejemplo los soundtrack de juegos retro sin mucha complejidad instrumental. El interés de la investigación esta en agregar el contexto a la generación de música, no crear música de muy alta calidad.

4 Solución propuesta

Nuestra solución para el problema consistiría en una red neuronal como LSTM que sea capaz de generar música, pero que además reciba un input que contenga 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 parecida 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 va a ser 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 el contexto para el cual queremos generar música.

5 Descripción de experimentos a realizar

5.1 Experimentación

Dado a que el proyecto tiene un enfoque investigativo, hay varias decisiones que en este momento no hemos tomado dado que requerimos experimentar con ellas:

1. Arquitectura del generador: En primer lugar vamos a tener que probar distintas arquitecturas y modelos para el generador de música, con el fin de poder seleccionar la más adecuada para esta tarea. Un buen punto de partida es considerar modelos como LSTM [10], o la versión LSTM-GAN, o también los modelos generativos de Google Magenta [9]. A partir de ahí, tendremos que variar la estructura para incluirle un input para el contexto y otras posibles modificaciones para que el modelo logre funcionar.

2. Formato del contexto: Actualmente la idea de contexto que estamos dando es un poco abstracta, y se va a requerir mayor experimentación para decidir de que 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: Debemos decidir que datos de música le entregaremos al modelo para su entrenamiento. Nuestra base es entregarle música en formato Midi al generador para entrenarlo, pero queda por definir que pistas específicamente le vamos a entregar(por ejemplo entregarle sólo de un género musical en específico). También se podría considerar entregar más datos a parte del Midi como la metadata de la canción.

5.2 Métricas de evaluación

Dado que nuestro objetivo busca generar nueva música en vez de recomendar otras ya existentes, las métricas tradicionales no nos serviran mucho. En cambio, hay 2 aspectos para los cuales queremos implementar métricas con el fin de asegurar que nuestro objetivo se este cumpliendo:

1. Musicalidad: Básicamente esta métrica mide que lo que estemos generando pueda ser considerado música y no simplemente ruido o sonidos sin relación. Este punto es muy importante ya que nos interesa estar generando 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 sino cualquier otra cosa.

2. Adecuación al contexto: Esta métrica mediría 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, mencionaré brevemente aquí otras ideas que tenemos para el proyecto para así poder tener feedback de cuál podría ser mejor considerando que nuestra intención es posiblemente crear algo 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 cuál 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, podriamos 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 imágen: Esta idea consiste en recomendar música que vaya acorde al estilo de una imágen, utilizando técnicas de extracción de estilo para obtener una representación de la imágen, y luego con ese input recomendar

música que la acompañe de buena manera. A diferencia de nuestro proyecto actual, aquí no estariamos generando música sino que solo recomendando de temas ya existentes, por lo que sería más fácil que el proyecto actual que queremos realizar. 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 si son dominios muy distintos.

- Generación de música para recomendación pero ahora con un contexto de texto en vez de imágenes. 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 linea 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.

Solicitamos respetuosamente feedback para estas ideas dado que nos gustaría mucho hacer un proyecto publicable, pero sabemos que el proyecto presentado en la propuesta puede ser muy difícil de realizar en este tiempo, por lo cuál nos serviría saber que tan buenas estan estas 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, Minho 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 the35th 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/nnnnnnn. nnnnnnn

[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)