Plan de Investigación
Cargo JTP Dedicación Exclusiva

Martin A. Miguel - LU 181/09 - N. Legajo 0167777

Introducción/Motivación

El presente plan de investigación propone el desarrollo de modelos de inteligencia artifical que, a partir de un estímulo musical simbólico, estimen para distintos momentos del mismo la certeza que tiene un oyente humano en su predicción de cómo continuará el mismo.

Las producciones artísticas o de entretenimiento deben mantener un balance entre la novedad y la familiaridad [1]. Esto es particularmente evidente en la música, donde el uso de repeticiones y estructuras de organización es un recurso común y fundamental. En este ámbito, un estímulo musical propone una estructura a partir de la cual un oyente puede generar predicciones sobre como continuará. Por otra parte, la música también tiene desvíos de la estructura propuesta a fin de generar sorpresa. Sin no estuviera establecida esta estructura, no habría sensación de sorpresa ya que el oyente no establecería predicciones en primer lugar. De esta forma, un estímulo musical debe balancear la novedad y la familiaridad, de forma que un oyente pueda generar predicciones que sean luego desafiadas [2, 3].

En este contexto, este trabajo propone desarrollar modelos de inteligencia artificial y neurociencia computacional que permitan estimar el grado de certeza de un oyente frente a un estímulo musical en distintos puntos del mismo. Estas herramientas son de utilidad para proveer información a compositores y enriquecer modelos de composición automática. Este análisis también puede funcionar como entrada adicional a otras tareas del campo de recuperación de información musical (*Music Information Retrieval* o *MIR*, en inglés), como ser segmentación o clasificación automática, así como sistemas de recomendación.

Desde las áreas de inteligencia artificial, procesamiento de señales y MIR se han desarrollado numerosos modelos computacionales que buscan comprender un estímulo musical. Ejemplo de tareas donde esto sucede es composición [4], clasificación de emociones [5] y segmentación automática [6] de estímulos musicales. No obstante, estos trabajos se enfocan en resolver la tarea en cuestión y no en reflejar la forma en que un oyente humano procesa la música. De esta forma, no son herramientas para estimar la certeza de un oyente.

IDyOM es un modelo estadístico que fue desarrollado para la tarea de estimación de certeza en un estímulo musical [7]. El mismo permite estimar la probabilidad de distintas continuaciones a un segmento de un estímulo musical. Para ello se basa en regularidades estadísticas aprendidas a partir de un cuerpo de datos. Dada la distribución de probabilidad de las continuaciones, la certeza del oyente se estima a partir la entropía de la misma. La distribución de probabilidad estimada por el modelo, así como la estimación de certeza, fueron contrastadas con datos obtenidos de oyentes humanos [8]. En los experimentos, los participantes debían, para un contexto musical dado, puntuar el nivel de incertidumbre que tenían para posibles continuaciones. También, luego del contexto, eran presentados con distintas continuaciones y debían reportar que tan inesperadas eran las mismas. La estimación de probabilidad de IDyOM para las continuaciones mostró correlaciones significativas con las medidas de sorpresa de los participantes ($r = 0.695$, $p < 0.01$). De igual manera, la estimación de entropía del modelo correlacionó significativamente con los reportes de incertidumbre de los participantes ($r = 0.466$, $p = 0.02$).

El funcionamiento de IDyOM se basa en cadenas de Markov de orden variable, donde la distribución de probabilidad de una continuación para un contexto musical se establece a partir de contabilizar cuantos contextos similares al evaluado están presentes en los datos de entrenamiento y poseen la continuación en cuestión. Este mecanismo de inferencia puede ser extendido considerando otros mecanismos de aprendizaje y resumen que son utilizados en la cognición humana. Se considera que la misma es jerárquica, manteniendo concepciones del

mundo de distintos niveles de abstracción [9]. Por ejemplo, en el procesamiento de secuencias, las personas mantienen distribuciones de transición (similares a las cadenas de Markov), pero también realizan procesos de agrupación de la secuencia en secciones relevantes. Asimismo, pueden abstraer patrones algebraicos abstractos que luego se instancian con valores específicos (por ejemplo, la secuencia abstracta AAB, que puede instanciarse en 112 o 225) y abstraer estas aún más en árboles de parseo [10].

Objetivos

Este plan de investigación propone desarrollar modelos para la estimación de la certeza a lo largo del tiempo que tendría un oyente humano sobre las continuaciones de un estímulo musical dado un contexto.

En primer lugar, se propone la extensión de las ideas propuestas en el modelo IDyOM [7] con nuevos mecanismos de inferencia jerárquica como los mencionados en la sección antecedentes [10].

En segundo lugar, se propone reutilizar modelos de composición automática basados en redes neuronales que aprenden recurrencias estadísticas en estímulos musicales para obtener estimadores de certeza. Esto puede realizarse a partir de métricas obtenidas de las activaciones de la red neuronal, así como reutilizando las capas inferiores de las mismas y agregando una nueva capa superior entrenada para esta nueva tarea (proceso conocido como *fine-tunning*).

En tercer lugar, se propone realizar un experimento para recolectar información de certeza en contextos musicales. Estos datos serán utilizados para evaluar los modelos así como para mejorar el proceso de entrenamiento al contar con mayor cantidad de datos.

Metodología de trabajo

Para extender las propuestas de IDyOM, se propone el uso de modelos de inferencia Bayesiana jerárquica que utilizan gramáticas para describir los datos de entrada [11]. Por un lado, el uso de inferencia bayesiana tiene ineherentemente una estimación de probabilidad y por lo tanto de certeza. Por el otro, el uso de gramáticas permite hacer descripciones de secuencias de datos basadas en árboles. Estas gramáticas, por ejemplo, permiten reflejar la forma jerárquica en la que se anota la música escrita en partituras, donde los elementos constituyentes (las notas) se agrupan de forma recursiva [12]. Esta técnica de modelado ha mostrado ser flexible para representar distintos tipos de datos (taxonomías de animales, ubicaciones espaciales, distancia entre colores) y obtener información relevante a partir de pocos datos [11]. Asimismo, el modelado del aprendizaje utilizando gramáticas se ha utilizado para reflejar características de procesamiento humano, como ser tiempo de respuesta o dificultad de la tarea [13].

Para obtener métricas a partir de modelos de redes neuronales, proponemos inspeccionar modelos de composición automática secuenciales basados en redes neuronales. Estos modelos son entrenados para, dado un contexto musical, estimar una distribución de probabilidad de continuaciones de forma que repliquen el dataset [14]. De esta forma, y al igual que en IDyOM, la certeza del oyente puede estimarse a partir de la entropía de esta distribución de probabilidad. Esta metodología fue utilizada con éxito en un trabajo previo para obtener un estimador de la certeza del pulso musical [15].

Para el experimento, es posible realizar tareas donde los participantes reportan su certeza respecto de la continuación de distintos contextos musicales [8]. Para tomar medidas durante el proceso de escucha sin inducir una pausa, es posible utilizar mediciones fisiológicas como pupilometría [16],

Descripción del grupo de investigación

El trabajo de investigación se realizará dentro del Laboratorio de Inteligencia Artificial Aplicada. El mismo se enfoca en la aplicación de técnicas de inteligencia artificial y aprendizaje automático a distintos problemas. En particular, se mantienen líneas de investigación donde se desarrollan modelos que buscan reflejar y predecir características de la cognición humana (como ser la trayectoria de movimientos oculares en la percepción de imágenes [17], estimación de riesgo de psicosis [18] o predictibilidad de palabras en un texto [19]).

Factibilidad

Para el desarrollo de los modelos de inferencia no es necesario equipo especializado. Para el trabajo con redes neuronales puede ser necesario, según si se deben entrenar modelos, computadoras equipadas con placas de vídeo, las cuales están disponibles en el laboratorio. Para la realización de experimento se requiere una sala donde conducirlos. El laboratorio cuenta con una sala designada para ello. Además cuenta con una cámara de seguimiento ocular en caso de recolectar datos de pupilometría.

Otros

Se espera publicar el trabajo de modelado utilizando modelado basado en gramáticas en la revista de investigación técnica en música *Journal of New Music Research*. Se espera publicar el trabajo realizado a partir de redes neuronales en la conferencia de la *International Society of Music Information Retrieval* (ISMIR). Los datos experimentales recolectados podrán ser publicados en la misma conferencia.

Algunas secciones de este plan pueden ser llevadas a cabo en conjunto con un estudiante, ya sea como tesis de grado o beca de investigación. Un ejemplo es la exploración sobre modelos de redes neuronales, donde, dado un modelo ya seleccionado y un conjunto de datos de evaluación establecido, la/el estudiante deberá familiarizarse con el modelo, agregar código para obtener métricas del mismo, realizar una evaluación de las mismas y presentar los resultados.

Postulante (Martin A. Miguel)	Profesor (Diego Fernandez Slezak)

Referencias

[1] D.E. Berlyne and de berlyne. *Aesthetics and Psychobiology*. Century psychology series. Appleton-Century-Crofts, 1971. ISBN 9780390086709. URL `https://books.google.com.ar/books?id=o5TWAAAAMAAJ`.

[2] David Huron and Elizabeth Hellmuth Margulis. *Musical expectancy and thrills*, chapter 21, pages 575–604. Oxford University Press, 2010.

[3] Peter Vuust, Martin J. Dietz, Maria Witek, and Morten L. Kringelbach. Now you hear it: a predictive coding model for understanding rhythmic incongruity. *Annals of the New York Academy of Sciences*, 1423(1):19–29, 2018. doi: https://doi.org/10.1111/nyas.13622. URL `https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/nyas.13622`.

[4] Jean-Pierre Briot, Gaëtan Hadjeres, and François-David Pachet. *Deep learning techniques for music generation*, volume 1. Springer, 2020.

[5] XHJS Downie, Cyril Laurier, and MBAF Ehmann. The 2007 mirex audio mood classification task: Lessons learned. In *Proc. 9th Int. Conf. Music Inf. Retrieval*, pages 462–467, 2008.

[6] Brian McFee, Oriol Nieto, Morwaread M Farbood, and Juan Pablo Bello. Evaluating hierarchical structure in music annotations. *Frontiers in psychology*, 8:1337, 2017.

[7] Marcus Thomas Pearce. *The construction and evaluation of statistical models of melodic structure in music perception and composition*. PhD thesis, City University London, 2005.

[8] Niels Chr. Hansen and Marcus T. Pearce. Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, 5, 2014. ISSN 1664-1078. doi: 10.3389/fpsyg.2014.01052. URL https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01052.

[9] Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127, Feb 2010. ISSN 1471-0048. doi: 10.1038/nrn2787. URL https://doi.org/10.1038/nrn2787.

[10] Stanislas Dehaene, Florent Meyniel, Catherine Wacongne, Liping Wang, and Christophe Pallier. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron*, 88(1):2–19, 2015.

[11] Joshua B Tenenbaum, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285, 2011.

[12] W Tecumseh Fitch. Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience*, 7:68, 2013. ISSN 1662-5137. doi: 10.3389/fnsys.2013.00068. URL https://www.frontiersin.org/article/10.3389/fnsys.2013.00068.

[13] P. Tano, S. Romano, M. Sigman, A. Salles, and S. Figueira. Learning is Compiling: Experience Shapes Concept Learning by Combining Primitives in a Language of Thought. *ArXiv e-prints*, May 2018.

[14] Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew M Dai, Matthew D Hoffman, Monica Dinculescu, and Douglas Eck. Music transformer. *arXiv preprint arXiv:1809.04281*, 2018.

[15] Nicolás Pironio, Diego Fernandez Slezak, and Martin A Miguel. Pulse clarity metrics developed from a deep learning beat tracking model. In *Proceedings of the 22nd International Society for Music Information Retrieval Conference*, pages 525–530, Online, November 2021. ISMIR. doi: 10.5281/zenodo.5625692. URL https://doi.org/10.5281/zenodo.5625692.

[16] Felicia Zhang and Lauren L Emberson. Using pupillometry to investigate predictive processes in infancy. *Infancy*, 25(6):758–780, 2020.

[17] Gaston Bujia, Melanie Sclar, Sebastian Vita, Guillermo Solovey, and Juan Esteban Kamienkowski. Modeling human visual search in natural scenes: A combined bayesian searcher and saliency map approach. *Frontiers in systems neuroscience*, 16, 2022.

[18] Gillinder Bedi, Facundo Carrillo, Guillermo A. Cecchi, Diego Fernández Slezak, Mariano Sigman, Natália B. Mota, Sidarta Ribeiro, Daniel C. Javitt, Mauro Copelli, and Cheryl M. Corcoran. Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophrenia*, 1(1):15030, Aug 2015. ISSN 2334-265X. doi: 10.1038/npjschz.2015.30. URL https://doi.org/10.1038/npjschz.2015.30.

[19] Bruno Bianchi, Gastón Bengolea Monzón, Luciana Ferrer, Diego Fernández Slezak, Diego E. Shalom, and Juan E. Kamienkowski. Human and computer estimations of predictability of words in written language. *Scientific Reports*, 10(1):4396, Mar 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-61353-z. URL https://doi.org/10.1038/s41598-020-61353-z.