

Críticos de Arte Artificiales

Juan Romero¹, Penousal Machado², Bill Manaris³, Antonino Santos¹, Amílcar Cardoso², and Marisa Santos¹

¹ RNASA Lab., Fac. de Informática, Universidade da Coruña, España
{jj, nino, mhyso}@udc.es

² CISUC- Centre for Informatics and Systems, Universidade de Coimbra, Portugal
{machado, amilcar}@dei.uc.pt

³ Computer Science Department, College of Charleston, USA
{manaris@cs.cofc.edu}

Abstract. En este artículo proponemos un marco de trabajo para el desarrollo de críticos de arte artificiales, consistente en una arquitectura y en una metodología de validación. La arquitectura incluye dos módulos: un extractor de características, que lleva a cabo un preprocesamiento de la pieza de arte, extrayendo diversas medidas y características; y un evaluador, que, basado en la salida del extractor de características, clasifica la pieza de arte de acuerdo a un criterio específico. La metodología de validación consta de varias etapas, que van desde la identificación de autor y estilo, hasta la integración del crítico de arte artificial en un entorno dinámico multiagente, que incluye agentes humanos. Usando la estructura propuesta, hemos desarrollado un crítico de arte artificial en el dominio musical, presentado en los resultados experimentales.

1 Introducción

La habilidad para generar piezas de arte es comúnmente asociada con la creatividad. Así, el desarrollo de sistemas computacionales que crean estas piezas pueden contribuir de manera significativa al estudio de la creatividad.

El proceso artístico depende altamente de la habilidad de llevar a cabo juicios estéticos, de inspirarse en el trabajo de otros artistas, y de actuar como crítico del trabajo de uno mismo. Como Boden afirma: “alguien que tiene una nueva idea debe ser capaz de evaluarla por él mismo” [1]. Modelar esta capacidad del artista es un paso importante, sino necesario, en la creación de un artista artificial. Después de todo, un artista es también, y sobre todo, un observador/oyente.

Esto contrasta con la mayoría de los sistemas computacionales que han sido desarrollados durante los últimos años¹. Típicamente, el rol del observador/oyente ha sido completamente desatendido; tales sistemas no tienen la habilidad de percibir las piezas de arte producidas por ellos (o por otros artistas), ni son capaces de llevar a cabo juicios estéticos. Por lo tanto, estos sistemas tienden a ser completamente ciegos/sordos al mundo exterior.

¹ Para una visión general de las aproximaciones computacionales a la composición musical ver, por ejemplo, [2].

En este artículo presentamos un marco de trabajo general para el desarrollo de críticos de arte artificiales (CAAs), es decir, sistemas capaces de “ver/escuchar” una obra de arte y llevar a cabo algún tipo de evaluación sobre la misma. Estamos principalmente interesados en CAAs que producen una valoración numérica de las piezas de arte, puesto que esto permite una fácil incorporación del CAA al sistema de generación de piezas de arte. Además, este tipo de valoración no es particular a un estilo de arte específico. Aunque particularmente adecuado a este tipo de CAAs, este marco de trabajo es lo suficientemente general para permitir el desarrollo de otros tipos de CAAs (por ejemplo, aquellos que producen una evaluación descriptiva de la obra de arte).

Este marco de trabajo, basado en un análisis de CAAs existentes, consiste en una *arquitectura* y en una *metodología de validación*.

La arquitectura propuesta consta de un *extractor de características* y un módulo *evaluador*. El extractor de características es responsable de la percepción de la pieza de arte, generando como salida un conjunto de medidas que reflejan sus características relevantes. Estas medidas sirven como entrada al evaluador, que evalúa la pieza de arte según un criterio específico o estético.

Una de las principales dificultades en el desarrollo de artistas computacionales, y más específicamente CAAs, es su validación. Para ayudar a solucionar este problema, proponemos una *metodología de validación multietapa*. La primera etapa de esta metodología permite la valoración objetiva y significativa de los CAAs, proporcionando una base sólida para su desarrollo. Las últimas etapas incorporan criterios más dinámicos, e incluyen probar los CAAs en una sociedad híbrida de humanos y agentes artificiales.

Probamos nuestras ideas mediante el desarrollo de un CAA en el dominio musical, y realizamos un conjunto de experimentos, que, aunque preliminares, dan resultados prometedores.

2 Marco de trabajo para el desarrollo de CAAs

Queremos proporcionar una base para la validación y desarrollo de CAAs que permita la integración de otros críticos y promueva la colaboración entre grupos para la creación de CAAs. El marco de trabajo global está basado en las siguientes características:

- Adaptabilidad – Los CAAs deberían adaptarse a un entorno cambiante. Esto significa replicar una característica comúnmente aceptada entre los críticos humanos: la evolución.
- Sociabilidad – Idealmente, los CAAs deberían poder ajustar su comportamiento según las demandas de la sociedad en la cual están integrados. Esto es, los CAAs deben ser capaces de desarrollarse en un entorno híbrido (un entorno que incorpora humanos y sistemas artificiales). Así, el CAA debe ser validado por una sociedad de “agentes” artificiales y humanos, de la misma forma que los críticos humanos son validados en sociedades puramente humanas [3].

- Generalidad – El CAA debería ser fácilmente adaptable a diferentes dominios; las tareas de dominio específico deberían ser llevadas a cabo por módulos especializados, permitiendo al sistema cambiar fácilmente de un dominio a otro.
- Independencia de representación – El CAA debería construir su propia representación interna de la pieza de arte, determinando su evaluación a partir de la misma; y solo debería tener acceso a la pieza de arte, no a otro tipo de nivel superior de representación de la misma.

2.1 Arquitectura

Teniendo en cuenta las características presentadas previamente, la arquitectura debe permitir el desarrollo de CAAs fácilmente adaptables a diferentes dominios, y que consideren las particularidades de los mismos. Por ejemplo, la forma de tratar con la música y con el arte visual es visiblemente distinta: mientras que la música sigue una secuencia temporal predeterminada, el arte se observa mediante un acceso más directo a la pieza de arte. De ahí la necesidad de dividir el sistema en varios módulos, siendo específicos en tareas particulares a un dominio, y permitiendo la generalidad de los módulos restantes.

Normalmente las piezas de arte contienen una gran cantidad de información. En arte visual, por ejemplo, incluso un cuadro relativamente pequeño consume una gran cantidad de memoria. Como se puede deducir a partir del análisis del estado del arte de sistemas adaptativos actuales (redes de neuronas, algoritmos genéticos...), tales cantidades de información no pueden ser manejadas razonablemente. Para abordar este problema, algunos investigadores recurren a la reducción del tamaño de las obras de arte que alimentan el sistema adaptativo (p.ej. [4]). Sin embargo, esta aproximación implica una importante pérdida de información y detalle, y los resultados experimentales son, típicamente, decepcionantes.

Creemos que hay una aproximación más adecuada, que consiste en preprocesar las piezas para extraer características relevantes, que pueden entonces ser usadas como entrada a la parte adaptativa del sistema. Esto reduce la cantidad de información que tiene que ser procesada.

La arquitectura propuesta incluye dos módulos: el *extractor de características* y el *evaluador*. Cada módulo tiene un propósito concreto y diferente. El extractor de características lleva a cabo un análisis de la pieza de arte y proporciona un conjunto de características relevantes al evaluador. El evaluador realiza una valoración de la pieza basado en las características extraídas previamente.

El extractor de características lleva a cabo dos tareas específicas: percepción y análisis. Durante la percepción, el sistema construye un tipo de representación interna de la pieza de arte. Después, en la tarea de análisis, esta representación es analizada y proporciona un conjunto de medidas relevantes. Mientras que esta separación entre percepción y análisis es principalmente conceptual, la idea es que, en una primera etapa, el extractor de características adquiera información sobre parámetros específicos al dominio que son analizados posteriormente.

La representación interna no está restringida a técnicas específicas, sino que puede ser de diferentes tipos: estadística, algorítmica, simbólica, no simbólica, etc.

El evaluador es un sistema adaptativo que toma como entrada la caracterización de la pieza de arte realizada por el extractor de características, y obtiene como salida una valoración de dicha pieza.

Este módulo debe adaptarse a diferentes tareas según la información de retroalimentación que se le proporciona. Dependiendo de la tarea, esta información indica la respuesta deseada o una evaluación de la actuación del CAA, el cual debe ajustar su comportamiento para maximizar su rendimiento.

Además, el módulo evaluador adaptativo puede dar información acerca de las características que son relevantes en la evaluación de una pieza de arte. Los pesos de una red de neuronas artificiales (RNA), por ejemplo, pueden mostrar qué características son las más significativas a la hora de criticar una pieza. También es posible encontrar el conjunto mínimo de características necesarias para una cierta tarea mediante pruebas de test al evaluador con diferentes conjuntos.

La arquitectura propuesta permite que la búsqueda de características relevantes y la evaluación sean independientes. Así, es posible incluir extractores de características y evaluadores de diferentes investigadores que permitan probar qué combinación de extractor y evaluador es la que caracteriza mejor a la pieza de arte. Ahora presentaremos una metodología de validación que fue diseñada para realizar un test estructurado del CAA desarrollado.

2.2 Metodología de validación

La validación de un CAA presenta principalmente dos dificultades: la subjetividad existente en la evaluación de las piezas de arte, y el hecho de que sean necesarios grandes conjuntos de entrenamiento para entrenar el módulo evaluador (cientos de piezas de arte evaluadas por humanos).

La respuesta a estas complicaciones es el uso de una *metodología de validación multietapa*. En cada nivel, el CAA se presenta con una tarea diferente. Se empieza con tareas en las que la exactitud de la salida puede ser objetivamente determinada, y que no requieren un conjunto de piezas de arte evaluadas por humanos. En la siguiente etapa las tareas requieren una mayor subjetividad y complejidad. En los primeros niveles la respuesta del CAA es estática; en el último nivel, sin embargo, el CAA debe adaptarse al entorno y cambiar su evaluación en el tiempo según el contexto que lo rodea.

Actualmente, consideramos tres niveles de validación: Identificación, Evaluación Estática y Evaluación Dinámica.

El *nivel de identificación* se ocupa de la valoración de la habilidad del CAA para reconocer el autor o el estilo de una determinada obra.

En la tarea de *identificación de autor* se presenta al CAA varias piezas de arte de diferentes autores. Su objetivo es determinar el autor de cada pieza. El módulo evaluador es entrenado con información de retroalimentación que le indica la tarea a realizar. Este tipo de validación es relativamente fácil de llevar a cabo, la compilación de instancias de entrenamiento es sencilla, y el test es totalmente objetivo. La principal dificultad que afecta a este nivel de prueba es la construcción de conjuntos de entrenamiento y de test representativos.

La tarea de *identificación de estilo* es similar a la anterior. La diferencia es que en este caso el CAA debe identificar el estilo de una obra. El entrenamiento y el test pueden ser realizados del mismo modo que en la etapa de identificación de autor. Este tipo de validación permite la comprobación de CAAs que pueden ser usados en una amplia variedad de tareas, tales como recuperación de imágenes y música, permitiendo búsquedas basadas en el estilo.

La principal dificultad de estas tareas depende de los artistas y estilos escogidos. Intentar discriminar entre artistas de la misma escuela puede ser más difícil que distinguir estilos radicalmente distintos. Sin embargo, discriminar entre artistas que tienen rasgos característicos (en el sentido usado en [5]) es más fácil que entre estilos relacionados. En el análisis de los resultados de los experimentos es importante tener en cuenta qué es razonable esperar. Por ejemplo, si el conjunto de test incluye obras atípicas, el CAA probablemente fallará. Esto no indica necesariamente un defecto del extractor de características o del evaluador, sino simplemente el hecho de que una pieza de arte es atípica.

Aunque de ámbito limitado, las tareas de identificación son útiles para determinar las capacidades del módulo extractor de características. Un fallo en estas pruebas puede indicar que el conjunto de características extraídas no es suficiente para discriminar entre autores o estilos, previniéndonos de esta manera desplazarnos a una tarea más compleja, que seguramente falle debido a la falta de información significativa. Además, un análisis de las características usadas por el evaluador en las tareas de identificación puede ayudar a determinar la importancia relativa de cada una de ellas. De hecho, se pueden desarrollar tests específicos para conocer el poder predictivo de cada medida o conjunto de medidas.

El segundo nivel de validación es la *evaluación estática*. El objetivo del CAA es determinar el valor estético de una serie de piezas de arte previamente evaluadas por humanos. Una de las mayores dificultades para desarrollar este test es la construcción de una base de datos representativa con obras evaluadas consistentemente.

Es importante darse cuenta de que el entrenamiento del CAA requiere no solo ejemplos positivos, sino también negativos. Irónicamente, es bastante difícil conseguir un conjunto representativo de “cosas incorrectas que hacer”.

Para crear el conjunto de entrenamiento, se puede recurrir a una herramienta de arte generativa. Esto produciría un alto número de piezas en una cantidad de tiempo razonable. Sin embargo, la consistencia de la evaluación depende mayormente de la disciplina del usuario. Adicionalmente, el conjunto será solo representativo de piezas típicamente creadas por una herramienta de arte generativa. Además, el grado de correlación entre las piezas creadas puede ser alto, haciendo la tarea del CAA más fácil.

Otra opción sería disminuir el ámbito de aplicación del CAA; esto es, crear un CAA que es capaz de evaluar la calidad estética dentro de un estilo bien definido. Esto da lugar a un paso de validación que está de alguna forma más cercano a la tarea de identificación de estilo y que es, por tanto, menos subjetiva. La diferencia está en que el CAA está valorando la distancia a un estilo dado en vez de intentar distinguir entre estilos.

El análisis de los resultados experimentales puede ser muy significativo; uno necesita estar seguro de que el CAA está llevando a cabo la tarea esperada y no explotando algún defecto del conjunto de entrenamiento. Por ejemplo en [6] los

autores entrenaban un sistema de reconocimiento de caras, que tenía un resultado sorprendentemente bueno. Sin embargo, un cuidadoso análisis de los resultados mostró que el sistema no estaba reconociendo las caras de las personas que aparecían en las imágenes, sino las oficinas en las cuales estaban tomadas las fotos.

La etapa de evaluación estática tiene muchas dificultades, tanto en la construcción del test como en el análisis de los resultados experimentales, pero es, sin embargo, necesaria para poder evaluar un CAA.

El último paso en la metodología es la *evaluación dinámica*. El valor de una obra depende del contexto cultural que la rodea. Así, el CAA debe ser “consciente” de este contexto, y ser capaz de adaptar su evaluación a los cambios que se dan en su entorno. Esto es, su comportamiento debe ser socialmente adecuado. Para desarrollar esta validación, se propone un modelo de sociedad llamado “Sociedad Híbrida” (SH). SH es un paradigma similar a la Vida Artificial, pero con “agentes” humanos en el mismo nivel que los artificiales. La sociedad híbrida explora la creación de sociedades igualitarias pobladas por seres humanos y artificiales en dominios artísticos (o sociales); como tal, SH es adecuada para validar el CAA de una forma natural y dinámica. En esta etapa, el éxito del CAA depende de la estimación de sus juicios por los otros miembros de la sociedad. Este tipo de test introduce una nueva dimensión social y dinámica a la validación, puesto que el valor de una obra varía con el tiempo, y depende de los agentes que componen la sociedad.

El problema de este nivel de validación es la necesidad de incorporar humanos en la experimentación. Así, los experimentos son difíciles de organizar, y existen limitaciones de tiempo. Además, la capacidad de adaptación de los críticos debe ser alta para poder adaptarse a un entorno dinámico y complejo. A pesar de las dificultades inherentes, estos críticos pueden ser evaluados y fácilmente integrados en la “sociedad de la información” como asistentes de usuarios o como parte de sistemas generales.

En los dos primeros niveles de validación es posible valorar la ejecución del feature extractor y del evaluador independientemente, puesto que la salida del feature extractor (junto con la información de retroalimentación), puede ser vista como una instancia de entrenamiento para el evaluador. En el tercer nivel, esto ya no es posible ya que la información de retroalimentación no refleja directamente la calidad de las obras, sino solamente una estimación de las acciones del CAA mediante la sociedad, que cambia dinámicamente con el tiempo.

La metodología de validación aquí presentada intenta encontrar un compromiso entre la validación humana y la automática. Conscientes de la dificultad de las tareas propuestas, es importante resaltar que para ciertas tareas solo se necesitan tener en cuenta algunos de los niveles de validación.

3 Experimentos: Resultados

Usando el marco de trabajo presentado, desarrollamos un CAA en el dominio musical y comprobamos su comportamiento mediante un conjunto de experimentos,

que se corresponden con el primer nivel de validación. La tarea presentada al CAA es discriminar entre obras musicales de dos compositores: Beethoven y Bach. El conjunto de piezas musicales consiste en 108 partituras de Bach (una colección de sonatas, preludios, fugas, fantasías, toccatas, conciertos, etc.) y 32 sonatas de piano de Beethoven.

Siguiendo la arquitectura propuesta, el sistema tiene dos módulos, el extractor de características y el evaluador adaptativo, que son descritos en las siguientes secciones.

3.1 Extractor de características

El extractor de características, descrito en [7], utiliza una colección de métricas, basadas en la Ley de Zipf [8], para extraer una serie de métricas a partir de piezas de músicas codificadas en formato MIDI.

Las distribuciones Zipf han sido descubiertas en un amplio rango de fenómenos, incluyendo la música. Por ejemplo, en [9] se presenta un estudio de 220 piezas de varios estilos de música (barroco, clásico, romántico, doce-tonos, jazz, rock, cadenas de ADN, y música aleatoria), descubriendo en ellas varias distribuciones Zipf.

En los experimentos se usan un total de 40 métricas, además del número de notas de la obra. Cada una de las 40 métricas produce dos números reales:

1. La *pendiente* de la línea de dirección de las frecuencias de evento, trazada en un formato log-log, rango-frecuencia; este número varía entre 0 y $-\infty$, con -1 denotando una distribución Zipf; y
2. La fuerza de la correlación lineal, R^2 , de la línea de dirección; ésta abarca desde el 0 al 1, siendo el 1 el que indica un ajuste perfecto.

Las métricas usadas en el extractor de características se pueden dividir en 3 tipos:

- Las *métricas globales* proporcionan información estadística de la pieza como un todo. Hay siete métricas de este tipo: tono, tono-relativo-a-octava, duración×tono, duración×tono-relativo-a-octava, intervalo melódico, intervalo harmónico e intervalo melódico-harmónico.
- Las *métricas estructurales* miden el equilibrio de órdenes más altos de cambio de tono. Actualmente, capturamos seis órdenes de cambio. Las métricas de primer-orden miden el equilibrio de los cambios en los intervalos melódicos. Las métricas de segundo-orden miden el equilibrio de los cambios entre los intervalos de primer-orden, y así el resto.
- Las *métricas fractales* miden la dimensión fractal de cada una de las métricas anteriores. Estas métricas aplican recursivamente una métrica dada en diferentes niveles de resolución dentro de una pieza. Mediante la subdivisión sucesiva de la pieza en partes, la carencia de equilibrio local puede ser expuesto. Como las otras métricas, las métricas fractales producen una pendiente y un valor del error cuadrático medio. La pendiente es equivalente a la dimensión fractal de la métrica dada. El proceso de particionamiento para cuando alcanzamos frases con menos de cinco notas.

3.2 Evaluador adaptativo

El evaluador adaptativo usado en los experimentos consiste en una *RNA feed-forward* (con alimentación hacia delante) con una capa oculta. Tras probar con diferentes arquitecturas de redes, elegimos una con 30 neuronas en la capa de entrada, 12 en la oculta y 2 en la de salida. Cada unidad de la capa de entrada se corresponde con cada uno de los valores generados por las métricas. Estos valores se normalizan en el intervalo $[-1, 1]$. Una salida (1, 0) indica que el autor de la partitura es Beethoven, mientras que (0, 1) indica que pertenece a Bach. El conjunto de entrenamiento usado en los experimentos contiene un 66% de las partituras (aleatoriamente seleccionadas) de cada compositor. El conjunto de test contiene las restantes.

En los experimentos preliminares, se descubrió que cuando no se incluía una partitura atípica en el conjunto de entrenamiento, la RNA fallaba al identificar a su autor. Este problema se solucionó al incluir esta partitura en el conjunto de entrenamiento. En la sección 3.3 se hace un análisis detallado de este problema.

Usamos el SNNS² para construir, entrenar y probar las RNAs. Los pesos son aleatoriamente inicializados con valores del intervalo $[-1, 1]$. La función de aprendizaje utilizada es la *back-propagation estándar* (propagación hacia atrás), con una tasa de aprendizaje de 0.1 y un momentum igual a 0. En los primeros tests usamos una RNA completamente conectada. El entrenamiento de la RNA se realiza con 30000 ciclos. Tras el entrenamiento la red es capaz de identificar correctamente todas las partituras de los conjuntos de entrenamiento y test. El error cuadrático medio (ECM) en el ciclo 30000 era 0.00003 en el conjunto de entrenamiento, y 0.00576 en el conjunto de test.

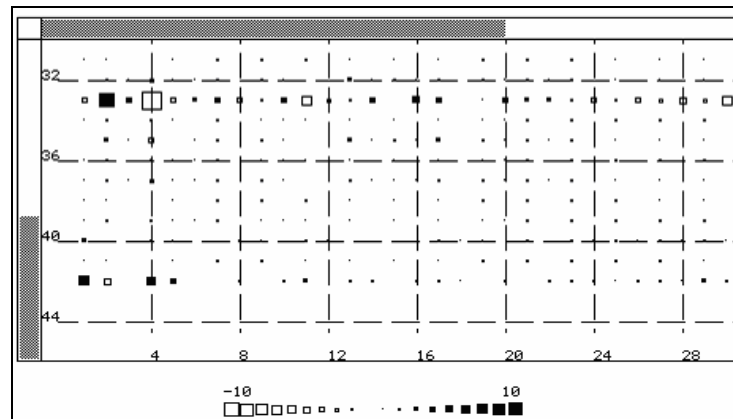


Fig. 1. Conexión entre los pesos de los elementos de procesamiento de la capa de entrada (eje x) y los de la capa oculta (eje y) de la RNA

² Stuttgart Neural Network Simulator (<http://www-ra.informatik.uni-tuebingen.de/SNNS/>).

Para discriminar las características más relevantes para la tarea de identificación, sumamos los pesos en valor absoluto de las conexiones de cada una de las neuronas de la capa de entrada con las neuronas de la capa oculta (Figure 1). Los valores más altos están asociados a las características más relevantes. Tras varias repeticiones del experimento, se observan similares distribuciones de pesos vinculados a las mismas características. Para comprobar esto se realizaron los siguientes tests.

Comenzamos sacando las ocho neuronas menos significativas. Los resultados obtenidos son similares a los anteriores. Tras 30000 ciclos la red identifica correctamente todas las partituras de los conjuntos de entrenamiento y test (ECMs de 0.00010 y 0.00356, respectivamente).

Después construimos una RNA con sólo 6 unidades de entrada, correspondientes con las características más relevantes. Los resultados muestran una ligera degradación del proceso. En el ciclo 10000, la RNA identifica correctamente todas las partituras del conjunto de entrenamiento y el 94% de las del conjunto de test (ECMs de 0.00391 y 0.11166, respectivamente). A partir de este punto, el error en el test se incrementa gradualmente, alcanzando un ECM de 0.12776 en el ciclo 70000, lo que indica que la RNA se ha sobreentrenado.

3.3 Análisis de los experimentos

Los resultados obtenidos muestran que las métricas basadas en Zipf combinadas con la RNA, son suficientes para la identificación de autor. Además se destaca un conjunto de seis características relevantes para discriminar entre los dos autores. Cabe mencionar que esto no significa necesariamente que sean las características más importantes en diferentes tareas o con diferentes autores.

Las Figuras 2.a y 2.b muestran los mapas de contorno en 3D de las obras de Bach y Beethoven. En estas visualizaciones, el eje x se corresponde con las 6 características más significativas; el eje y se corresponde con una pieza musical (1 a 32); y el eje z se corresponde con el valor absoluto de la característica. Analizando estas figuras se ve que la pendiente del tono-relativo-a-octava es uno de los factores clave para discernir las obras de Bach y Beethoven. Como se muestra, las piezas de Bach exhiben una distribución casi-Zipfiana de los tonos de la escala-12-cromática (la pendiente media es -1.1629; std 0.2809), mientras que las sonatas de Beethoven tienden a ser más distribuidas uniformemente (pendiente media -0.8343; std 0.2188).

La Figura 2.b permite la identificación de una pieza, la Sonata para Piano no. 20 de Beethoven, que no se ajusta al contorno de Beethoven. La pendiente del tono-relativo-a-octava para esta partitura es -1.7472. Esta pieza es atípica con respecto a las otras obras de Beethoven y a las características usadas. Como tal, para ser clasificada correctamente, esta pieza tuvo que ser incluida en el conjunto de entrenamiento. La inclusión de esta instancia de entrenamiento fuerza a la RNA a basar su valoración en un conjunto más amplio de características, lo que, aunque provoca un entrenamiento más lento, fomenta la robustez y generalización de la RNA.

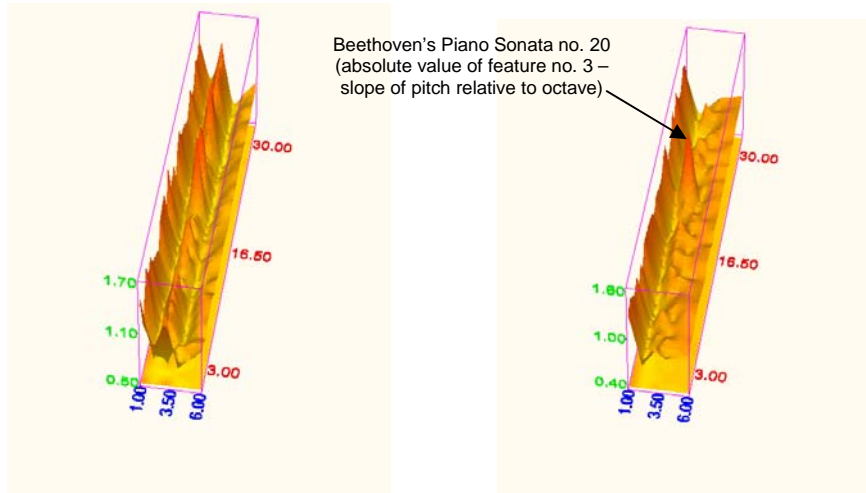


Fig. 3.a. Bach- Mapa de contorno de 6 características sobre 32 piezas de Bach (BWV 500-531)

Fig. 3.b. Beethoven- Mapa de contorno de 6 características sobre 32 piezas de Beethoven (piano sonatas 1-32)

Una característica secundaria distintiva entre los dos contornos musicales es el R^2 del tono-relativo-a-octava. Las obras de Bach tienden a producir una línea de dirección más dispersa (media R^2 0.6612; std 0.0787), mientras que las de Beethoven tienden a tener una línea de dirección más ajustada (media R^2 0.8017; std 0.0571). Aunque éstas son las dos características más destacadas, es necesario tener en cuenta el conjunto de las seis características para lograr la correcta identificación de todas las partituras.

4 Conclusiones y futuros trabajos

Proponemos un marco de trabajo genérico para el desarrollo de críticos de arte artificiales, basado en el análisis del estado del arte actual en el área, y en la experiencia adquirida en el desarrollo de sistemas previos. Este marco de trabajo incluye una arquitectura y una metodología de validación. Para permitir una fácil adaptación a diferentes dominios, la arquitectura propuesta separa los componentes genéricos del dominio de los específicos. Además también establece un límite entre los módulos estáticos y adaptativos. La validación de CAAs es una tarea compleja, por lo que la metodología multinivel permite probarlos estructuradamente y compararlos con diferentes aproximaciones.

Siguiendo con el marco propuesto, implementamos un CAA y probamos su eficiencia en la identificación de autor. El extractor de características obtiene una serie de métricas basadas en Zipf, que sirven como entrada al evaluador

adaptativo, implementado mediante una RNA. Una vez entrenada es capaz de reconocer todas las instancias de los conjuntos de entrenamiento y test, mostrando la eficacia de las métricas basadas en Zipf para la identificación de autor. Éste es el primer uso de dichas métricas para la atribución de autoría en música. Un análisis de la RNA nos permitió identificar el conjunto de características que son más importantes para la discriminación de autores. La identificación de estas características puede ser útil desde una perspectiva musical, dando una idea de la caracterización de los estilos de los autores. La naturaleza modular de la arquitectura permite una fácil integración de características adicionales y la adaptación a otros dominios.

Actualmente estamos desarrollando numerosos experimentos en el dominio del arte visual y musical, que incluye la discriminación entre más autores, con un conjunto de características mayor. Una posibilidad interesante consiste en explorar si el principio de mínimo esfuerzo de Zipf podría ser usado, en un nivel más alto, para evaluar la eficacia y la “naturalidad” de una sociedad arbitraria igualitaria de CAAs, mediante el examen de diversos aspectos de la interacción social entre agentes.

El marco de trabajo aquí descrito no está restringido a dominios artísticos. Puede ser usado en algunos dominios que incluyen (a) la creación de una hipótesis (diseño, solución, etc.), y (b) el refinamiento iterativo de aquellas hipótesis basadas en la estética, restricciones, y otros atributos cuantificables. Tales dominios incluyen el desarrollo de software, matemáticas, ingeniería, y arquitectura. Por ejemplo, las métricas Zipf ya han sido usadas para evaluar software, diseño arquitectónico y otros sistemas complejos [10; 11]. Adicionalmente, la aplicación a otras áreas tales como imágenes basadas en contenido, y búsqueda y recuperación de música, también parece viable.

La investigación en el área de los críticos de arte artificiales todavía está en una etapa embrionaria. El sistema propuesto pretende proporcionar una base común para el desarrollo y la validación de críticos de arte artificiales, y promocionar la colaboración entre investigadores en esta área.

Agradecimientos

Expresamos nuestra gratitud a Robert Davis por sus diversos análisis estadísticos; Charles McCormick, Tarsem Purewal, Dallas Vaughan y Christopher Wagner por contribuir en el desarrollo de las métricas basadas en Zipf.

References

1. M. A. Boden. *The Creative Mind: Myths and Mechanisms*. London, Cardinal. 1990.

2. G. Papadopoulos and G. A. Wiggins. AI Methods for Algorithmic Composition: A Survey, A Critical View, and Future Prospects. *Proceedings of the AISB'99 Symposium on Musical Creativity*, 1999.
3. A. Pazos, A. Santos, B. Arcay, J. Dorado, J. Romero, and J. Rodríguez. An Application Framework for Building Evolutionary Computer Systems in Music. *Leonardo*, 36(1), 2003.
4. S. Baluja, D. Pomerleau, and T. Jochem. Towards Automated Artificial Evolution for Computer-Generated Images. In *Connection Science* 6, No. 2, pp. 325–354. 1994.
5. David Cope. *Experiments in Musical Intelligence*. Madison, WI: A-R Editions, 1996.
6. A. Teller and M. Veloso. Algorithm evolution for face recognition: What makes a picture difficult. In *Proceedings of the International Conference on Evolutionary Computation*, IEEE Press, 1995.
7. B. Manaris, T. Purewal and C. McCormick. Progress Towards Recognizing and Classifying Beautiful Music with Computers-MIDI-Encoded Music and the Zipf-Mandelbrot Law. In *Proceedings of IEEE SoutheastCon 2002*, Columbia, SC, pp. 52–57, 2002.
8. G. K. Zipf. *Human Behavior and the Principle of Least Effort*. New York: Hafner Publishing Company, 1949.
9. Manaris, B., Vaughan, D., Wagner, C., Romero, J., and Davis, R.: Evolutionary Music and the Zipf-Mandelbrot Law: Developing Fitness Functions for Pleasant Music. In: *Lecture Notes in Computer Science, Applications of Evolutionary Computing*. LNCS 2611, Springer-Verlag, pp. 522–534, 2003.
10. M. Shooman and A. Laemmel. Statistical Theory of Computer Programs. In *Proceedings of IEEE Computer Conference*, pp. 511–517, Oct. 1977.
11. N.A. Salinger and B. J. West. A Universal Rule for the Distribution of Sizes. *Environment and Planning*, B(26), pp. 909–923, 1999.