

Seminario de Aprendizaje Topológico 2025-2

Can Transformers do enumerate geometry?

por

B. Hashemi, R. G. Corominas y A. Giacchetto

Miguel Evangelista

19 de febrero de 2025

Contenidos

Abstract

Introducción

Estructuras cuanticas de Airy

Métodos

Resultados

Representacion de los datos

Modelo

Conclusión

Referencias

Abstract

El trabajo explora la aplicación de modelos Transformer para la geometría enumerativa computacional, centrado específicamente en el cálculo de los números de intersección de clase ψ en el espacio de moduli de las curvas.

Abstract

El trabajo explora la aplicación de modelos Transformer para la geometría enumerativa computacional, centrado específicamente en el cálculo de los números de intersección de clase ψ en el espacio de moduli de las curvas.

La contribución clave de este trabajo es la introducción de una nueva función de activación denominada *Dynamic Range Activator (DRA)*, diseñada para capturar el comportamiento recursivo de los números de intersección.

Abstract

Más allá del cálculo de los números de intersección, el estudio investiga la capacidad de los Transformers para modelar las restricciones de Virasoro basada en datos.

Abstract

Más allá del cálculo de los números de intersección, el estudio investiga la capacidad de los Transformers para modelar las restricciones de Virasoro basada en datos.

Además, mediante pruebas de hipótesis abductivas, se proporciona evidencia de que el modelo comprende varios valores que aparecen en la asíntota de gran género de los números de intersección de clases ψ .

Contenidos

Abstract

Introducción

Estructuras cuanticas de Airy

Métodos

Resultados

Representacion de los datos

Modelo

Conclusión

Referencias

Geometría enumerativa

La geometría enumerativa es una rama de la geometría algebraica se ocupa de contar objetos geométricos o de encontrar invariantes que satisfagan determinadas condiciones geométricas [Mum83, Kat06].

Geometría enumerativa

La geometría enumerativa es una rama de la geometría algebraica se ocupa de contar objetos geométricos o de encontrar invariantes que satisfagan determinadas condiciones geométricas [Mum83, Kat06].

Una pregunta clásica de la geometría enumerativa es: ¿Cuántas cónicas pasan por cinco puntos dados del plano?

Geometría enumerativa

La geometría enumerativa es una rama de la geometría algebraica se ocupa de contar objetos geométricos o de encontrar invariantes que satisfagan determinadas condiciones geométricas [Mum83, Kat06].

Una pregunta clásica de la geometría enumerativa es: ¿Cuántas cónicas pasan por cinco puntos dados del plano?

Otra más actual es: ¿Cuántas curvas racionales de grado uno hay en un *quintic threefold*?

La resolución de estos problemas a menudo se reduce a calcular los números de intersección.

La resolución de estos problemas a menudo se reduce a calcular los números de intersección.

Los números de intersección son fundamentales para calcular los volúmenes de Weil-Petersson, los números de Hurwitz, las invariantes de Gromov-Witten y las amplitudes de la gravedad cuántica. Un ejemplo de estos objetos son los números de intersección de clase ψ .

Métodos recursivos tradicionales como la jerarquía KdV [Wit91, Kon92] tienen complejidades computacionales que crecen del orden $O(g! \cdot 2n)$, donde g es el género y n el número de puntos marcados, haciendo que los cálculos sean inviables incluso para curvas sencillas.

Métodos recursivos tradicionales como la jerarquía KdV [Wit91, Kon92] tienen complejidades computacionales que crecen del orden $O(g! \cdot 2n)$, donde g es el género y n el número de puntos marcados, haciendo que los cálculos sean inviables incluso para curvas sencillas.

Es bien sabido que los Transformers tienen dificultades para aprender con patrones periódicos, por no hablar de los problemas recursivos [ZHU20].

Geometría enumerativa y IA

Se introduce una función de activación no lineal, el *Activador de Rango Dinámico (DRA)*, que es particularmente adecuado para el aprendizaje de funciones recursivas.

Geometría enumerativa y IA

Se introduce una función de activación no lineal, el *Activador de Rango Dinámico (DRA)*, que es particularmente adecuado para el aprendizaje de funciones recursivas.

Se desarrolla *DynamicFormer*, un modelo Transformer diseñado para predecir números de intersección de clase ψ dados los datos iniciales de la estructura cuántica de Airy [KS17].

Geometría enumerativa y IA

Se introduce una función de activación no lineal, el *Activador de Rango Dinámico (DRA)*, que es particularmente adecuado para el aprendizaje de funciones recursivas.

Se desarrolla *DynamicFormer*, un modelo Transformer diseñado para predecir números de intersección de clase ψ dados los datos iniciales de la estructura cuántica de Airy [KS17].

Con lo anterior se pretende introducir un nuevo paradigma para el uso de Transformadores en geometría algebraica.

Contenidos

Abstract

Introducción

Estructuras cuanticas de Airy

Métodos

Resultados

Representacion de los datos

Modelo

Conclusión

Referencias

Estructuras cuanticas de Airy

El enfoque conceptualmente más sencillo para calcular los números de intersección de clase ψ implica el uso de restricciones de Virasoro reformuladas en el lenguaje de las estructuras cuánticas de Airy [KS17].

Estructuras cuanticas de Airy

El enfoque conceptualmente más sencillo para calcular los números de intersección de clase ψ implica el uso de restricciones de Virasoro reformuladas en el lenguaje de las estructuras cuánticas de Airy [KS17].

Estas últimas son una reformulación algebraica de la recursividad topológica [EO07].

Estructuras cuánticas de Airy

Una estructura cuántica de Airy sobre un espacio vectorial complejo V , generada por $(x_i)_{i \in I}$, es una familia de operadores diferenciales $(L_i)_{i \in I}$ en el álgebra de Weyl sobre V de la forma:

$$L_i = \hbar \partial_i - \hbar^2 \sum_{a,b \in I} \left(\frac{1}{2} A_{i,a,b} x^a x^b + B_{i,a}^b x^a \partial_b + \frac{1}{2} C_i^{a,b} \partial_a \partial_b \right) - \hbar^2 D_i, \quad (2.1)$$

Estructuras cuánticas de Airy

Una estructura cuántica de Airy sobre un espacio vectorial complejo V , generada por $(x_i)_{i \in I}$, es una familia de operadores diferenciales $(L_i)_{i \in I}$ en el álgebra de Weyl sobre V de la forma:

$$L_i = \hbar \partial_i - \hbar^2 \sum_{a,b \in I} \left(\frac{1}{2} A_{i,a,b} x^a x^b + B_{i,a}^b x^a \partial_b + \frac{1}{2} C_i^{a,b} \partial_a \partial_b \right) - \hbar^2 D_i, \quad (2.1)$$

Formando una subálgebra del álgebra de Weyl, donde $\delta_i = \delta_x^i$ y $A_{i,a,b} = A_{i,b,a}$, $B_{i,a}^b = B_{i,a}^b$, $C_i^{a,b} = C_i^{b,a}$, y D_i , son escalares indexados por elementos en I . \hbar es un parámetro formal que lleva la cuenta del género.

Estructuras cuánticas de Airy

Dada una estructura cuántica de Airy, se puede encontrar unívocamente una función formal Z sobre V , llamada función de partición, que es anulada por los operadores diferenciales $(L_i)_{i \in I}$.

Estructuras cuánticas de Airy

Dada una estructura cuántica de Airy, se puede encontrar unívocamente una función formal Z sobre V , llamada función de partición, que es anulada por los operadores diferenciales $(L_i)_{i \in I}$.

Más precisamente, se cumple el siguiente teorema.

Estructuras cuánticas de Airy

Dada una estructura cuántica de Airy, se puede encontrar unívocamente una función formal Z sobre V , llamada función de partición, que es anulada por los operadores diferenciales $(L_i)_{i \in I}$.

Más precisamente, se cumple el siguiente teorema.

Theorem 2.1. *If $(L_i)_{i \in I}$ is a quantum Airy structure on V , there exists a unique formal function Z of the form*

$$Z(x; \hbar) = \exp \left(\sum_{\substack{g \geq 0 \\ n > 0}} \frac{\hbar^{2g-2+n}}{n!} \sum_{d_1, \dots, d_n \in I} F_{g; d_1, \dots, d_n} x^{d_1} \dots x^{d_n} \right) \quad (2.2)$$

with $F_{0; d_1} = F_{0; d_1, d_2} = 0$, $F_{g; d_1, \dots, d_n}$ symmetric in $d_1, \dots, d_n \in I$, and such that

$$L_i Z = 0 \quad \forall i \in I. \quad (2.3)$$

Estructuras cuanticas de Airy

Los elementos $F_{g,n} = \sum_{d_1, \dots, d_n \in I} F_{g; d_1, \dots, d_n} x^{d_1, \dots, d_n}$ son tensores simétricos de rango n sobre V^* .

Estructuras cuanticas de Airy

Los elementos $F_{g,n} = \sum_{d_1, \dots, d_n \in I} F_{g; d_1, \dots, d_n} x^{d_1, \dots, d_n}$ son tensores simétricos de rango n sobre V^* .

La colección de coeficientes (A, B, C, D) también puede considerarse como tensores, y a menudo nos referiremos a ellos como los datos iniciales de la estructura cuántica de Airy.

Estructuras cuánticas de Airy

Para una estructura cuántica de Airy dada, los coeficientes $F_{g;d_1,\dots,d_n}$, llamados amplitudes de la estructura cuántica de Airy, se determinan recursivamente sobre la característica de Euler $2g - 2 + n$.

Estructuras cuánticas de Airy

Para una estructura cuántica de Airy dada, los coeficientes $F_{g;d_1,\dots,d_n}$, llamados amplitudes de la estructura cuántica de Airy, se determinan recursivamente sobre la característica de Euler $2g - 2 + n$.

Para los valores bajos de g y n , los casos base vienen dados por

$$F_{0;i,j,k} := A_{i,j,k} \text{ y } F_{1;i} := D_i,$$

donde $A_{i,j,k}$ y D_i forman parte de los datos de la estructura cuántica de Airy.

Estructuras cuanticas de Airy

Para valores más altos de g y n que satisfagan $2g - 2 + n > 0$, la fórmula de recursión, conocida como recursión topológica, viene dada por

Estructuras cuanticas de Airy

Para valores más altos de g y n que satisfagan $2g - 2 + n > 0$, la fórmula de recursión, conocida como recursión topológica, viene dada por

$$F_{g;d_1,d_2,\dots,d_n} = \sum_{m=2}^n \sum_{a \in I} B_{d_1,d_m}^a F_{g;a,d_2,\dots,\widehat{d_m},\dots,d_n} + \frac{1}{2} \sum_{a,b \in I} C_{d_1}^{a,b} \left(F_{g-1;a,b,d_2,\dots,d_n} + \sum_{\substack{g_1+g_2=g \\ I_1 \sqcup I_2 = \{d_2,\dots,d_n\}}} F_{g_1;a,I_1} F_{g_2;b,I_2} \right) \quad (2.4)$$

Estructuras cuanticas de Airy

En general, el cálculo de las amplitudes tiene una complejidad computacional de $O(g! \cdot 2n)$, lo que hace problemático el cálculo de las amplitudes en géneros altos.

Estructuras cuanticas de Airy

En general, el cálculo de las amplitudes tiene una complejidad computacional de $O(g! \cdot 2n)$, lo que hace problemático el cálculo de las amplitudes en géneros altos.

Para espacios vectoriales de dimensión alta V , encontrar una solución resulta cada vez menos práctico, si no imposible.

Estructuras cuanticas de Airy

El ejemplo central es el asociado a los números de intersección de clase ψ .

Estructuras cuanticas de Airy

El ejemplo central es el asociado a los números de intersección de clase ψ .

En este caso, el espacio vectorial subyacente está generado por vectores indexados por $I = \mathbb{N}$, los enteros no negativos.

Estructuras cuanticas de Airy

El ejemplo central es el asociado a los números de intersección de clase ψ .

En este caso, el espacio vectorial subyacente está generado por vectores indexados por $I = \mathbb{N}$, los enteros no negativos.

Los operadores diferenciales $(L_i)_{i \geq 0}$ vienen determinados por los tensores

$$\begin{aligned} A_{i,j,k} &:= \delta_{i,j,k}, \\ B_{i,j}^k &:= \delta_{i+j,k+1} \frac{(2k+1)!!}{(2i+1)!!(2j-1)!!}, \\ C_i^{j,k} &:= \delta_{i,j+k+2} \frac{(2j+1)!!(2k+1)!!}{(2i+1)!!}, \\ D_i &:= \frac{\delta_{i,1}}{24} \end{aligned} \tag{2.5}$$

Estructuras cuanticas de Airy

Los operadores resultantes, tras desplazar los índices como $L_i = L_{i-1}$, forman una representación del álgebra de Virasoro:

$$[L_i, L_j] = h^2(i - j)L_{i+j}$$

para todo $i \geq -1$.

Estructuras cuanticas de Airy

Los operadores resultantes, tras desplazar los índices como $L_i = L_{i-1}$, forman una representación del álgebra de Virasoro:

$$[L_i, L_j] = h^2(i - j)L_{i+j}$$

para todo $i \geq -1$.

Sorprendentemente, las amplitudes asociadas coinciden con los números de intersección de clase ψ .

Estructuras cuanticas de Airy

Los operadores resultantes, tras desplazar los índices como $L_i = L_{i-1}$, forman una representación del álgebra de Virasoro:

$$[L_i, L_j] = h^2(i - j)L_{i+j}$$

para todo $i \geq -1$.

Sorprendentemente, las amplitudes asociadas coinciden con los números de intersección de clase ψ .

$$F_{g;d_1,\dots,d_n} = \begin{cases} \langle \tau_{d_1} \cdots \tau_{d_n} \rangle_{g,n} & \text{if } d_1 + \cdots + d_n = d_{g,n} , \\ 0 & \text{otherwise.} \end{cases} \quad (2.6)$$

Estructuras cuanticas de Airy

Nos centramos en la elección específica de los datos de la estructura cuántica de Airy dados por la ecuación (2.5), calculando así los números de intersección de clase ψ .

Estructuras cuánticas de Airy

Nos centramos en la elección específica de los datos de la estructura cuántica de Airy dados por la ecuación (2.5), calculando así los números de intersección de clase ψ .

Para facilitar la notación, denotaremos una partición genérica de $d_{g,n}$ de longitud n como $d = (d_1, \cdot, d_n)$. El número de intersección de clase ψ asociado se denominará $\langle d \rangle_{g,n} := \langle \tau d_1 \cdots d_n \rangle_{g,n}$.

Contenidos

Abstract

Introducción

Estructuras cuanticas de Airy

Métodos

Resultados

Representacion de los datos

Modelo

Conclusión

Referencias

Dynamic Range Activator

Como se ha visto a lo largo de la exposición, la naturaleza recursiva de las funciones en la geometría enumerativa, introduce una gran complejidad en la modulación y aproximación de dichas funciones.

Dynamic Range Activator

Como se ha visto a lo largo de la exposición, la naturaleza recursiva de las funciones en la geometría enumerativa, introduce una gran complejidad en la modulación y aproximación de dichas funciones.

Por lo tanto, para capturar dicho comportamiento y realizar una inferencia adecuada, se introduce el Activador de Rango Dinámico (DRA), y el modelo Transformer llamado DynamicFormer.

Dynamic Range Activator

Como se ha visto a lo largo de la exposición, la naturaleza recursiva de las funciones en la geometría enumerativa, introduce una gran complejidad en la modulación y aproximación de dichas funciones.

Por lo tanto, para capturar dicho comportamiento y realizar una inferencia adecuada, se introduce el Activador de Rango Dinámico (DRA), y el modelo Transformer llamado DynamicFormer.

El DRA está diseñado para manejar las propiedades de crecimiento recursivo y factorial inherentes a los problemas enumerativos con una sobrecarga computacional mínima y puede integrarse en las redes neuronales existentes sin requerir cambios arquitectónicos significativos.

Dynamic Range Activator

El DRA integra los componentes armónicos e hiperbólicos de la siguiente manera,

Dynamic Range Activator

El DRA integra los componentes armónicos e hiperbólicos de la siguiente manera,

$$\text{DRA}(x) := x + a \sin^2\left(\frac{x}{b}\right) + c \cos(bx) + d \tanh(bx), \quad (3.1)$$

donde a, b, c, d son parámetros aprendibles.

Dynamic Range Activator

El DRA integra los componentes armónicos e hiperbólicos de la siguiente manera,

$$\text{DRA}(x) := x + a \sin^2\left(\frac{x}{b}\right) + c \cos(bx) + d \tanh(bx), \quad (3.1)$$

donde a, b, c, d son parámetros aprendibles.

Permite a la función modelar simultáneamente datos periódicos (a través del seno y el coseno) y respuestas de crecimiento rápido (a través de la tangente hiperbólica).

Dynamic Range Activator

DRA se inspira en la función no lineal Snake (Ziyin et al., 2020).

Dynamic Range Activator

DRA se inspira en la función no lineal Snake (Ziyin et al., 2020).

Sin embargo, Snake solo ofrece una modulación sinusoidal añadida a un término lineal que, si bien proporciona una transformación no lineal básica, carece de la flexibilidad adicional para los efectos de amortiguación o amplificación rápida que pueden proporcionar las tangentes hiperbólicas.

Dynamic Range Activator

El primer experimento es entrenar una red neuronal totalmente conectada con dos capas ocultas formadas por 64 y 32 neuronas y con funciones de activación ReLU, Tanh, Gated Linear Unit (GLU) y Sanke. Además, se compara con una red Kolmogorov-Arnold (KAN). Los resultados se muestran en la Figura 1.

Dynamic Range Activator

El primer experimento es entrenar una red neuronal totalmente conectada con dos capas ocultas formadas por 64 y 32 neuronas y con funciones de activación ReLU, Tanh, Gated Linear Unit (GLU) y Sanke. Además, se compara con una red Kolmogorov-Arnold (KAN). Los resultados se muestran en la Figura 1.

Las funciones de activación antes mencionadas, son incapaces de capturar la naturaleza recursiva de la función subyacente en un tiempo de entrenamiento finito.

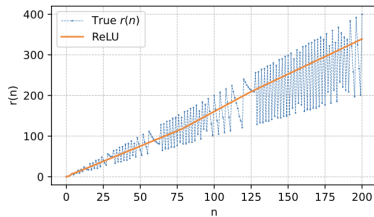
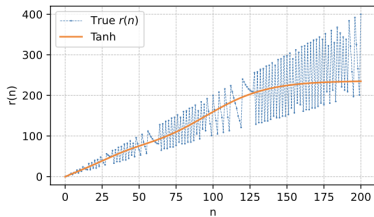
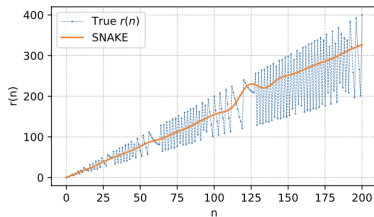
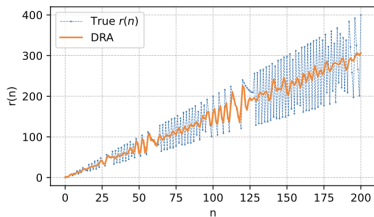
Dynamic Range Activator

El primer experimento es entrenar una red neuronal totalmente conectada con dos capas ocultas formadas por 64 y 32 neuronas y con funciones de activación ReLU, Tanh, Gated Linear Unit (GLU) y Sanke. Además, se compara con una red Kolmogorov-Arnold (KAN). Los resultados se muestran en la Figura 1.

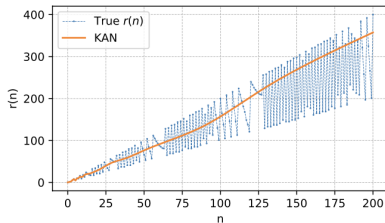
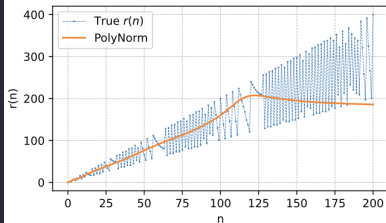
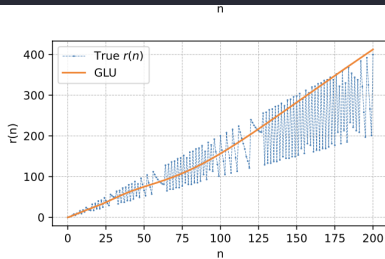
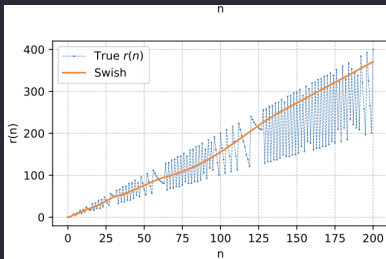
Las funciones de activación antes mencionadas, son incapaces de capturar la naturaleza recursiva de la función subyacente en un tiempo de entrenamiento finito.

Por el contrario, DRA muestra una mejor capacidad para capturar las fluctuaciones correctas.

Dynamic Range Activator



Dynamic Range Activator



Contenidos

Abstract

Introducción

Estructuras cuanticas de Airy

Métodos

Resultados

Representacion de los datos

Modelo

Conclusión

Referencias

Resultados

Para evaluar el rendimiento, consideramos dos configuraciones:
Dentro de la distribución (ID) y fuera de la distribución (OOD).
Utilizamos R^2 para comparar los resultados y evaluar la tarea de
regresión de números de intersección.

Resultados

Para evaluar el rendimiento, consideramos dos configuraciones: Dentro de la distribución (ID) y fuera de la distribución (OOD). Utilizamos R^2 para comparar los resultados y evaluar la tarea de regresión de números de intersección.

Para estimar la incertidumbre en las predicciones, utilizamos Predicción Conforme (CP) (Vovk et al., 1999), una técnica probabilística que genera intervalos de predicción sin asumir una distribución específica de los datos.

Coeficiente de determinación

La métrica R^2 (*coeficiente de determinación*), mide qué tan bien un modelo de regresión explica la variabilidad de la variable dependiente en función de las variables independientes.

Coeficiente de determinación

La métrica R^2 (*coeficiente de determinación*), mide qué tan bien un modelo de regresión explica la variabilidad de la variable dependiente en función de las variables independientes.

Se define como:

$$R^2 = 1 - \frac{SS_{\text{residual}}}{SS_{\text{total}}}$$

donde:

- ▶ SS_{residual} es la suma de los errores cuadráticos entre los valores reales y los predichos.
- ▶ SS_{total} es la suma de los errores cuadráticos entre los valores reales y su media.

Coeficiente de determinación

Interpretación

- ▶ $R^2 = 1$: El modelo explica toda la variabilidad de los datos.
- ▶ $R^2 = 0$: El modelo no explica ninguna variabilidad mejor que la media de los datos.
- ▶ $R^2 < 0$: El modelo es peor que simplemente usar la media como predicción.

Configuración ID

En la configuración ID, se examinan datos con los mismos géneros que los datos de entrenamiento, es decir, $g_{ID} = [1, 13]$, pero con un número diferente y desconocido de puntos marcados $n_{ID} \in [35, 11]$.

Configuración ID

En la configuración ID, se examinan datos con los mismos géneros que los datos de entrenamiento, es decir, $g_{ID} = [1, 13]$, pero con un número diferente y desconocido de puntos marcados $n_{ID} \in [35, 11]$.

En esta configuración, el R^2 , la anchura conforme (CW) se muestran en la siguiente tabla

Configuración ID

(g, n)	$R^2 \uparrow$	Coverage	CW
(1, 35)	99.8	90.35	1.03
(2, 33)	99.6	83.60	0.84
(3, 31)	99.9	74.79	0.76
(4, 29)	98.7	95.66	1.11
(5, 27)	99.1	92.66	1.03
(6, 25)	99.3	91.18	0.80
(7, 23)	99.1	93.05	0.68
(8, 21)	99.8	90.88	0.76
(9, 19)	99.9	96.71	0.91
(10, 17)	99.9	90.01	1.04
(11, 15)	99.8	91.97	0.87
(12, 13)	99.6	89.08	1.30
(13, 11)	99.9	95.90	0.97

(a) R^2 and conformal uncertainty estimation results with $\alpha = 0.1$ (90% target coverage) in the ID setting.

Configuración OOD

En la configuración OOD, se examinan datos con un género superior al de los datos de entrenamiento, concretamente $g_{OOD} = [14, 15, 16, 17]$, y un número de puntos marcados $n_{OOD} \in [1, 9]$.

Configuración OOD

En la configuración OOD, se examinan datos con un género superior al de los datos de entrenamiento, concretamente $g_{OOD} = [14, 15, 16, 17]$, y un número de puntos marcados $n_{OOD} \in [1, 9]$.

En esta configuración, el R^2 , la anchura conforme (CW) se muestran en la siguiente tabla

Configuración OOD

(g, n)	$R^2 \uparrow$	Coverage	CW
$(14, [1, 9])$	99.6	93.82	0.93
$(15, [1, 7])$	95.9	84.27	0.91
$(16, [1, 5])$	94.1	89.60	3.55
$(17, [1, 3])$	93.8	95.27	8.30

(b) R^2 and conformal uncertainty estimation results with $\alpha = 0.1$ (90% target coverage) in the OOD setting.

Contenidos

Abstract

Introducción

Estructuras cuanticas de Airy

Métodos

Resultados

Representacion de los datos

Modelo

Conclusión

Referencias

Representación de los datos

Los datos de entrada durante el entrenamiento consistieron en los tensores dispersos B y C de la ecuación (2.5), el género g , el número de puntos n , y las particiones $d = (d_1, \dots, d_n)$ de $d_{g,n}$.

Representación de los datos

Los datos de entrada durante el entrenamiento consistieron en los tensores dispersos B y C de la ecuación (2.5), el género g , el número de puntos n , y las particiones $d = (d_1, \dots, d_n)$ de $d_{g,n}$.

No se incluyeron A y D como entrada, ya que son condiciones iniciales fijas para todos los números de intersección de clase ψ .

Representación de los datos

Los datos de entrada durante el entrenamiento consistieron en los tensores dispersos B y C de la ecuación (2.5), el género g , el número de puntos n , y las particiones $d = (d_1, \dots, d_n)$ de $d_{g,n}$.

No se incluyeron A y D como entrada, ya que son condiciones iniciales fijas para todos los números de intersección de clase ψ .

Consideraremos valores de g y n tales que la dimensión $d_{g,n}$ nunca superará $d_{\max} = 100$, los datos iniciales consistirán sólo en un número finito de datos.

Representación de los datos

Los datos iniciales durante el entrenamiento se representaron de la siguiente manera.

- Datos iniciales de B :

Un tensor $B_k \in \mathbf{R}^{d_{g,n} \times d_{g,n} \times d_{g,n}}$. Dicho tensor une la intersección i, j del mismo género (ver la ecuación (2.4)).

Representacion de los datos

Los datos iniciales durante el entrenamiento se representaron de la siguiente manera.

► Datos iniciales de B :

Un tensor $B_k \in \mathbf{R}^{d_{g,n} \times d_{g,n} \times d_{g,n}}$. Dicho tensor una la intersección i, j del mismo género (ver la ecuación (2.4)).

Dado que se trata de un tensor disperso con $d_{g,n} \leq d_{max} = 100$, optamos por representar sus componentes distintos de cero en el formato COO. En el formato COO, B se representa como un conjunto de 4 tuplas, cada una de las cuales contiene los índices y el valor distinto de cero correspondiente:

$$B = \{(i, j, k, B_{i,j}^k) | B_{i,j}^k \neq 0\}$$

donde $i, j, k \in [0, d_{g,n}]$, y $B_{i,j}^k$ son las componentes no nulas del tensor. Con esta eleccion de d_{max} la longitud máxima de B es aproximadamente 1500.

Representación de los datos

► Datos iniciales de C :

Un tensor $C_i^{j,k} \in \mathbf{R}^{d_{g,n} \times d_{g,n} \times d_{g,n}}$. Dicho tensor une números de intersección de distintos géneros (ver la ecuación (2.4)).

Representacion de los datos

► Datos iniciales de C :

Un tensor $C_i^{j,k} \in \mathbf{R}^{d_{g,n} \times d_{g,n} \times d_{g,n}}$. Dicho tensor une números de intersección de distintos géneros (ver la ecuación (2.4)).

De nuevo, dado que se trata de un tensor disperso con $d_{g,n} \leq d_{max} = 100$, optamos por representar sus componentes distintos de cero en el formato COO. En el formato COO, C se representa como un conjunto de 4 tuplas, cada una de las cuales contiene los índices y el valor distinto de cero correspondiente:

$$B = \{(i, j, k, C_i^{j,k}) | C_i^{j,k} \neq 0\}$$

donde $i, j, k \in [0, d_{g,n}]$, y $C_i^{j,k}$ son las componentes no nulas del tensor. Con esta eleccion de d_{max} la longitud máxima de B es aproximadamente 1500.



Representación de los datos

En esta configuración, se excluyen los datos iniciales de C . Esto había sido observado por Aggarwal (Aggarwal, 2021), que los términos C contribuyen cuadráticamente, mientras que el término B contribuye linealmente y tiene un efecto más fuerte en el cálculo de las intersecciones de clase ψ .

Representación de los datos

► Particiones:

Los números de intersección $\langle d \rangle_{g,n}$ están etiquetados por particiones $d = (d_1, \dots, d_n) \in \mathbb{N}^n$ de $d_{g,n}$ de longitud n .

Representación de los datos

► Particiones:

Los números de intersección $\langle d \rangle_{g,n}$ están etiquetados por particiones $d = (d_1, \dots, d_n) \in N^n$ de $d_{g,n}$ de longitud n .

Una característica importante de los números de intersección (y más en general, de las amplitudes calculadas a partir de estructuras cuánticas de Airy) es que son invariantes bajo permutación de los índices, es decir, de los elementos de la partición d .

Representación de los datos

► Particiones:

Los números de intersección $\langle d \rangle_{g,n}$ están etiquetados por particiones $d = (d_1, \dots, d_n) \in \mathbb{N}^n$ de $d_{g,n}$ de longitud n .

Una característica importante de los números de intersección (y más en general, de las amplitudes calculadas a partir de estructuras cuánticas de Airy) es que son invariantes bajo permutación de los índices, es decir, de los elementos de la partición d .

Por ejemplo, dada una partición $d = (3, 0, 0)$, tenemos

$$\langle 3, 0, 0 \rangle_{1,3} = \langle 0, 3, 0 \rangle_{1,3} = \langle 0, 0, 3 \rangle_{1,3}.$$

Representacion de los datos

Entrenamos el modelo con datos hasta el género $g = 13$ y luego le encargamos que predijera los números de intersección para los géneros $g = 14, 15, 16$ y 17 .

Contenidos

Abstract

Introducción

Estructuras cuanticas de Airy

Métodos

Resultados

Representacion de los datos

Modelo

Conclusión

Referencias

Los Transformers pueden manejar eficazmente este tipo de estructura presentada anteriormente, debido a sus mecanismos de atención enmascarada, que son expertos en la captura de la estructura dispersa de B .

Los Transformers pueden manejar eficazmente este tipo de estructura presentada anteriormente, debido a sus mecanismos de atención enmascarada, que son expertos en la captura de la estructura dispersa de B .

Por otra parte, la entrada d es invariante de permutación, y los Transformers se adaptan naturalmente a esta propiedad a través de su mecanismo de auto-atención que trata los elementos de entrada simétricamente sin plantear ningún sesgo de ordenación.

Snake MLP		D
$R^2 \uparrow$	CW \downarrow	R^2
69.3	8.44	76



Modelo

La base de DynamicFormer, consiste en dos Transformadores modificados: uno discreto, que actúa sobre d , y otro continuo, que actúa sobre el tensor B .

La base de DynamicFormer, consiste en dos Transformadores modificados: uno discreto, que actúa sobre d , y otro continuo, que actúa sobre el tensor B .

Las posiciones de los elementos en B , representados en formato COO, corresponden a combinaciones particulares de índices que calculan los números de intersección.

Modelo

Finalmente toda la información incorporada es agregada por la capa Principal Neighbourhood Aggregation (PNA) (Corso et al., 2020), agrupando la información de las modalidades (B, d) .

Finalmente toda la información incorporada es agregada por la capa Principal Neighbourhood Aggregation (PNA) (Corso et al., 2020), agrupando la información de las modalidades (B, d) .

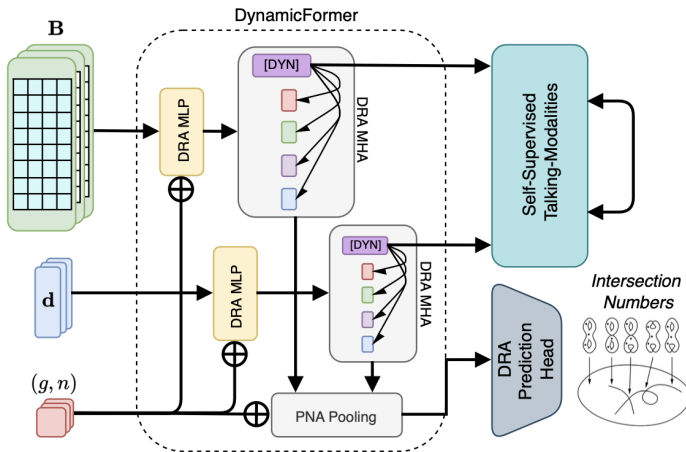
Una vez que la información de estas dos ramas, para cada muestra, se modula con sus correspondientes atributos posicionales de género y número de puntos marcados, (g, n) , se mapea a las salidas, es decir, números de intersección de clase ψ , mediante una cabezal MLP.

Finalmente toda la información incorporada es agregada por la capa Principal Neighbourhood Aggregation (PNA) (Corso et al., 2020), agrupando la información de las modalidades (B, d) .

Una vez que la información de estas dos ramas, para cada muestra, se modula con sus correspondientes atributos posicionales de género y número de puntos marcados, (g, n) , se mapea a las salidas, es decir, números de intersección de clase ψ , mediante una cabezal MLP.

DRA es la función de activación no lineal utilizada a lo largo de estos bloques.

Modelo



Modelo

Ilustración de DynamicFormer. Procesa dos modalidades principales de entrada:

- ▶ El dato B de la estructura cuántica de Airy como secuencia ordenada y las particiones d como conjunto invariante de permutaciones.
- ▶ El género g y el número de puntos n se incorporan como propiedades de entrada, moduladas con las modalidades principales en varias etapas.

Todas las capas, incluidos los bloques de Atención Multicabezal (MHA), utilizan la función de activación no lineal Activador de Rango Dinámico (DRA).

Modelo

Ilustración de DynamicFormer. Procesa dos modalidades principales de entrada:

- ▶ El dato B de la estructura cuántica de Airy como secuencia ordenada y las particiones d como conjunto invariante de permutaciones.
- ▶ El género g y el número de puntos n se incorporan como propiedades de entrada, moduladas con las modalidades principales en varias etapas.

Todas las capas, incluidos los bloques de Atención Multicabezal (MHA), utilizan la función de activación no lineal Activador de Rango Dinámico (DRA).

El cabezal de predicción DRA es un MLP de 2 capas que predice números de intersección de clase ψ en escala logarítmica.

Contenidos

Abstract

Introducción

Estructuras cuanticas de Airy

Métodos

Resultados

Representacion de los datos

Modelo

Conclusión

Referencias

Conclusión

El trabajo exploró si los Transformers pueden aprender y aplicar conceptos de geometría enumerativa y recursividad topológica, enfocándose en el cálculo de números de intersección de clase ψ .

Conclusión

El trabajo exploró si los Transformers pueden aprender y aplicar conceptos de geometría enumerativa y recursividad topológica, enfocándose en el cálculo de números de intersección de clase ψ .

Para ello, se introdujo DynamicFormer, un modelo basado en Transformers diseñado para predecir estos valores en un entorno de disparo cero. Un hallazgo clave fue que captar la naturaleza recursiva de las intersecciones representa un desafío significativo para estos modelos.

Conclusión

Para mejorar su rendimiento, se propuso una nueva función de activación, Dynamic Range Activator (DRA), que optimiza la precisión en mapas recursivos.

Conclusión

Para mejorar su rendimiento, se propuso una nueva función de activación, Dynamic Range Activator (DRA), que optimiza la precisión en mapas recursivos.

Además, se enfatizó la importancia de estimar la incertidumbre en las predicciones, incorporando técnicas de Conformal Prediction para hacer los resultados más fiables y explicables en el contexto del descubrimiento matemático.

Contenidos

Abstract

Introducción

Estructuras cuanticas de Airy

Métodos

Resultados

Representacion de los datos

Modelo

Conclusión

Referencias

Recurso Adicionales

- ▶ Github
- ▶ Colab

Referencias



Bertrand Eynard and Nicolas Orantin.

Invariants of algebraic curves and topological expansion.
Commun. Num. Theor. Phys., 1:347–452, 2007.



Sheldon Katz.

Enumerative Geometry and String Theory.
Student Mathematical Library. American Mathematical
Society, United States, 2006.



M. Kontsevich.

Intersection theory on the moduli space of curves and the
matrix Airy function.
Commun. Math. Phys., 147:1–23, 1992.



Maxim Kontsevich and Yan Soibelman.

*Airy structures and symplectic geometry of topological
recursion*.
2017.



David Mumford.