



INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN
LABORATORIO DE CIENCIA DE DATOS

GENERACIÓN DE CASCADAS ATMOSFÉRICAS USANDO
REDES GENERATIVAS ADVERSARIAS

T E S I S

PARA OBTENER EL TÍTULO DE:

MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA:

JOSÉ DE JESÚS DANIEL AGUIRRE ARZATE

TUTORES:

DR. RICARDO MENCHACA MÉNDEZ

DR. LUKAS NELLEN FILLA

México, CDMX
30 de mayo de 2022



Declaración de Autoría

I, José de Jesús Daniel AGUIRRE ARZATE, declare that this thesis titled, «Generación de cascadas atmosféricas usando redes generativas adversarias» and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

«If I have seen further it is by standing on the shoulders of Giants.»

Issac Newton

INSTITUTO POLITÉCNICO NACIONAL

Resumen

CIC

Centro de Investigación en Computación

Maestría en Ciencias de la Computación

Generación de cascadas atmosféricas usando redes generativas adversarias

por José de Jesús Daniel AGUIRRE ARZATE

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Agradecimientos

The acknowledgments and the people to thank go here, don't forget to include your project advisor...

Índice general

Declaración de Autoría	III
Resumen	VII
Agradecimientos	IX
1. Introduccion	1
1.1. Introducción	1
1.2. Planteamiento del problema	2
1.3. Objetivo de la tesis	2
1.4. Delimitación del tema	2
1.5. Organización de la tesis	3
2. Estado del arte	5
2.1. Antecedentes	5
2.2. Cascadas Atmosféricas	6
2.2.1. Propiedades de las cascadas atmosféricas	7
2.2.2. Métodos de detección	7
2.3. Análisis de las cascadas atmosféricas	8
2.4. Simulaciones	9
2.4.1. Estrategia para simulaciones de EAS	10
2.4.2. Problemática de las simulaciones	10
2.5. Modelos generativos	11
2.5.1. Aplicación de modelos generativos	11
3. Modelos generativos	13
3.1. Introducción	13
3.1.1. Fundamento teórico	13

3.2. Líneas de investigación	14
3.3. Modelos generativos	14
3.3.1. Conceptos básicos	15
3.3.2. Modelos basados en la verosimilitud	15
3.3.3. Modelos implícitos	17
GANS	17
WGAN	18
Conditional GAN	18
3.3.4. Aprendizaje de representación	18
4. Método experimental	21
4.1. Recolección y manejo de datos	21
4.2. Algoritmos	22
4.3. Desarrollo del método	22
4.3.1. Mapa del método experimental	22
5. Resultados	23
5.1. Dsicusion de resultados obtenidos	23
6. Conlusiones	25
6.1. Conclusiones puntuales obtenidas	25
6.2. Aplicacion y extension generadas del trabajo	25
6.3. Trabajos futuros	25
7. LaTeX Review	27
7.1. Welcome and Thank You	27
7.2. Learning L ^A T _E X	27
7.2.1. A (not so short) Introduction to L ^A T _E X	29
7.2.2. A Short Math Guide for L ^A T _E X	29
7.2.3. Common L ^A T _E X Math Symbols	29
7.2.4. L ^A T _E X on a Mac	30
7.3. Getting Started with this Template	30
7.3.1. About this Template	30
7.4. What this Template Includes	30

7.4.1. Folders	30
7.4.2. Files	31
7.5. Filling in Your Information in the <code>main.tex</code> File	32
7.6. The <code>main.tex</code> File Explained	32
7.7. Thesis Features and Conventions	33
7.7.1. Printing Format	33
7.7.2. Using US Letter Paper	34
7.7.3. References	34
A Note on bibtex	34
7.7.4. Tables	35
7.7.5. Figures	35
7.7.6. Typesetting mathematics	36
7.8. Sectioning and Subsectioning	37
7.9. In Closing	37
A. Frequently Asked Questions	39
A.1. How do I change the colors of links?	39
Bibliografía	41

Índice de figuras

2.1. Mecanismos de interacción	6
2.2. Distribucion longitudinal	7
2.3. Arreglo de detectores	8
2.4. Espectro de energía	9
2.5. Arquitectura ASICO	10
3.1. Modelo	16
3.2. Componentes de una GAN	17
7.1. An Electron	36

List of Algorithms

1. Algoritmo de entrenamiento de la GAN 22

Índice de cuadros

7.1. The effects of treatments X and Y on the four groups studied.	35
--	----

Índice de Abreviaturas

HEP	H igh E nergy P hysics
GEANT	G eoemtry A ND T racking
CORSIKA	C Osmic R ay S IMulations for K Ascade
GAN	G enerative A dversarial N etworks
DESY	D eutsches E lektronen- S ynchrotron
HAWC	H igh- A ltitude W ater C herenkov
SNOLAB	S udbury N eutrino O bservatory L ABoratory
CERN	C onseil E uropéen pour la R echerche N ucléaire
HL-LHC	H igh L uminosity L arge H adron C ollider
LAr	L iquid- A rgon
EAS	E xtensive A ir S hower
ASICO	A ir shower S imulation and C orrelation

Constantes Físicas

Speed of Light $c_0 = 2.997\,924\,58 \times 10^8 \text{ m s}^{-1}$ (exact)

Índice de Símbolos

eV	energy	J
a	distance	m
P	power	W (J s^{-1})
ω	angular frequency	rad

For/Dedicated to/To my...

Capítulo 1

Introduccion

En este capítulo se muestra la principal motivación para el desarrollo de esta tesis, así como una breve perspectiva de la problemática general. Como capítulo introductorio este contiene la introducción, el planteamiento del problema, el objetivo, las fronteras del estudio y la estructura del escrito.

1.1. Introducción

En el área de la física de altas energías (HEP) las técnicas de aprendizaje máquina siempre han estado presentes. Debido a la sorprendente efectividad de técnicas modernas del aprendizaje profundo, se comenzó a adaptar y desarrollar estos métodos en todos los rubros del campo. Algunas de las aplicaciones van desde los enfoques que se tienen en la parte experimental, la fenomenología o en el análisis teórico de los eventos.

En los experimentos más importantes del campo, el tratamiento y análisis de datos es una tarea fundamental. Técnicas como, árboles de decisión, máquinas de soporte vectorial, algoritmos genéticos, entre otras, fallan cuando la dimensionalidad de los datos aumenta. Como referencia de la alta dimensionalidad que hay, en el gran colisionador de hadrones (LHC), las colisiones ocurren con una frecuencia aproximada de 40Mhz, además cada colisión genera un gran número de partículas. Por ejemplo, en el caso particular del LHC, se tiene alrededor de $O(10^8)$ sensores para la detección de partículas.

Debido a que las observaciones son fundamentalmente probabilísticas se tiene un modelo estadístico que describe la probabilidad de observar un evento dado los parámetros de una teoría. Dado que el modelo de los datos experimentales no se conoce explícitamente, la alta dimensionalidad y a los grandes volúmenes de datos generan un problema. Sin embargo, se tiene acceso a muestras de datos generados por simuladores estocásticos que modelan la física de las interacciones.

Herramientas como PYTHIA, HERWING, GEANT, CORSIKA, se les denomina como simuladores Monte Carlo, los cuales cumplen con dos necesidades básicas. La primera, es aproximar el modelo estadístico que genera cada evento, la segunda, es generar una base de datos de simulaciones realistas.

Debido a que los simuladores describen las interacciones de las partículas con la materia, son computacionalmente demandantes y se llevan gran parte del presupuesto computacional de las colaboraciones experimentales [3], de modo que el desarrollo de simuladores rápidos es esencial.

Simuladores como GEANT y CORSIKA generan una excelente descripción de interacciones hadrónicas, sin embargo, son lentos para simular eventos de altas energías. De ahí la razón por la cual en los últimos años ha surgido un gran interés por usar modelos generativos para acelerar simuladores y tal vez llegar a usar estos métodos directamente en datos

generados por colisiones reales y hacer ajustes al momento.

El presente trabajo está fundamentado en el desarrollo de algoritmos de aprendizaje máquina profundo, específicamente el uso de modelos generativos implícitos como arquitecturas adversarias (GAN) para la generación de cascadas atmosféricas, debido a la necesidad de generar simulaciones precisas de una manera más rápida, a consecuencia de futuros requerimientos computacionales en las colaboraciones experimentales, por ejemplo, la actualización de alta luminosidad del LHC (HL-LHC)[3].

1.2. Planteamiento del problema

El planteamiento se fundamenta en lo siguiente:

Será posible diseñar un método que utilice redes neuronales generativas que logre simular cascadas atmosféricas precisas y así reducir el tiempo computacional de la generación de simulaciones mediante métodos tradicionales.

1.3. Objetivo de la tesis

Objetivo principal:

- Diseñar y implementar una red generativa adversaria para generar cascadas atmosféricas acordes a simulaciones de detectores de rayos cósmicos.

Objetivos particulares:

- I. Obtener datos de simulaciones acordes a cascadas atmosféricas.
- II. Diseñar y entrenar una red generativa adversaria para la simulación de cascadas atmosféricas.
- III. Ajustar la dimensión del vector latente del modelo y comparar resultados.
- IV. Obtener vectores latentes promedio y registrar como afecta a la cascada saliente.
- V. Cambiar de modelo o arquitectura y repetir.

1.4. Delimitación del tema

El presente trabajo se limita a utilizar al menos dos arquitecturas generativas en forma funcional para así lograr generar respuestas de detectores a una cascada atmosférica extensa. Enfatizando la rapidez de generación de nuevas muestras.

Cabe destacar que el trabajo se enfoca en la parte electromagnética de una cascada atmosférica, pero no se limita al análisis de otro tipo de cascadas de partículas o en su defecto, simulaciones de otro tipo de fenómenos físicos que involucran una multitud de procesos probabilísticos. Aunque no se pueda asegurar que el método presentado sea el mejor para la acelerar simulaciones, próximos estudios deben de intentar desenredar los parámetros latentes y probar la efectividad de otro tipo de arquitecturas.

Es necesario subrayar que este trabajo sólo utilizará arquitecturas adversarias (GAN) para acelerar simulaciones.

1.5. Organización de la tesis

El **Capítulo 1** presenta una breve y concisa introducción a la problemática principal, el planteamiento del problema, el objetivo principal y objetivos particulares del trabajo.

El **Capítulo 2** muestra el estado del arte, así como una breve explicación fenomenológica de las cascadas atmosféricas y la correspondiente operación básica de simuladores modernos. También se introducen las problemáticas principales relacionadas a los simuladores Monte Carlo.

En el **Capítulo 3** se introduce el fundamento teórico de los modelos generativos así como conceptos fundamentales del aprendizaje no supervisado. Asimismo se hace especial énfasis en las arquitecturas generativas y la teoría que las respalda.

El **Capítulo 4** presenta la metodología experimental que este trabajo sigue para lograr los objetivos propuestos.

En el **Capítulo 5 y 6** presentan los resultados obtenidos junto con una discusión de ellos y las conclusiones a las que este trabajo llegó.

Capítulo 2

Estado del arte

El presente capítulo proporciona el estado del arte mediante la revisión de conceptos y trabajos referentes a cascadas atmosféricas, detectores de partículas, simuladores y modelos generativos para establecer el fundamento del desarrollo de este trabajo.

2.1. Antecedentes

El problema de reducir el costo computacional que colaboraciones experimentales dedican a simulaciones en la física de altas energías (HEP), ha tenido mucha atención en los últimos años. Por esa razón, estudiar técnicas de aprendizaje máquina aplicadas a procesos experimentales, se incluyeron como un área estratégica de inversiones iniciales [3], para enfrentar futuros desafíos que actualizaciones como HL-LHC presentarán en los próximos años.

El aprendizaje máquina siempre ha estado presente en los flujos de trabajo de experimentos en HEP, sin embargo, técnicas modernas de aprendizaje profundo han comenzado a introducirse a procesos de análisis. En específico, se ha visto que los modelos generativos permiten acelerar los tiempos de generación de simulaciones, debido a lo anterior, es de suma importancia estudiar estas técnicas para así poder agregarlas a la próxima generación de simuladores, GEANTV, CORSIKA8, etc.

Algunos investigadores que han comenzado a usar modelos generativos para acelerar simulaciones son, Paganini [11, 12], Erdmann [6, 7], Carminati [1, 2], Glombitza [9, 5]. Paganini *et al*, desarrollan una arquitectura llamada CaloGAN basada en redes generativas adversarias para acelerar simulaciones de cascadas de partículas en calorímetros LAr y así lograr generar cascadas electromagnéticas tridimensionales con una reducción de tiempo computacional cinco órdenes de magnitud menor de lo que le toma a GEANT4.

Erdmann *et al*, usan una arquitectura WGAN para mejorar la estabilidad del entrenamiento y así poder reconstruir propiedades de la cascada simulada. El modelo se condiciona a un parámetro físico y logran generar simulaciones de un arreglo de calorímetros.

Carminati *et al*, presentan una red generativa adversaria convolucional tridimensional para generar la deposición energética de partículas en calorímetros de alta granularidad. Además, este trabajo es parte del proyecto GEANTV.

Glombitza usa redes convolucionales para reconstruir el máximo y la energía de una cascada atmosférica usando simulaciones de un arreglo de detectores, generadas con CORSIKA. Cabe señalar que el trabajo es parte de la colaboración Pierre Auger.

En ambos casos Paganini y Erdmann, proporcionan un modelo generativo capaz de generar simulaciones de cascadas de partículas acordes a algunos simuladores Monte Carlo

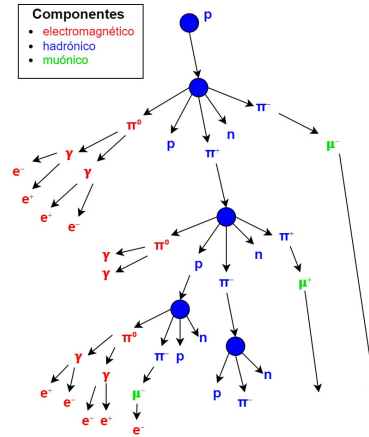


FIGURA 2.1: Mecanismo de la interacción de una partícula primaria con un componente atmosférico.

existentes. Consecuentemente, estos resultados muestran la posible efectividad de los modelos generativos para reducir los costos computacionales en los actuales flujos de trabajo experimentales en HEP.

2.2. Cascadas Atmosféricas

Problemas como conocer la naturaleza de la masa, la naturaleza de la antimateria, la dimensionalidad del espacio o lograr la unificación de las fuerzas fundamentales; se tienen que abordar tanto de los enfoques teórico como experimental, para así intentar responder a preguntas fundamentales como, ¿Cuántas partículas hay?, ¿Cuáles son sus propiedades y cómo interactúan?.

Estudiar la fenomenología de los rayos cósmicos es de vital importancia para poder responder las preguntas anteriores. En consecuencia laboratorios como DESY, SNOLAB, CERN, en conjunto con experimentos como el observatorio Pier Auger o el observatorio HAWC, entre otros; representan el actual estado del arte del campo de la física de altas energías. Por ejemplo, las partículas generadas en las colisiones del LHC viven por fracciones de segundo, lo cual hace que encontrar indicadores de que una partícula fue creada a pesar de nunca ser detectada, sea una tarea artesanal.

Los **rayos cósmicos** son partículas aceleradas en la galaxia o en objetos astrofísicos extragalácticos que al impactar con la atmósfera terrestre generan **cascadas de partículas**. El consenso general es que rayos cósmicos debajo de energías de $3 \cdot 10^6 \text{ GeV}$ son acelerados en los remanentes de supernovas galácticas y para energías mayores no se tiene una clara idea de que es lo que acelera a estas partículas, lo único que se tiene claro es de que dichas partículas tienen un origen extragaláctico. A modo de comparación, estas fuentes aceleran partículas en tres órdenes de magnitud mayor que el equivalente energético del LHC.

Cuando la energía de los rayos cósmicos sobrepasa significativamente 1000 GeV , estos tienen que ser estudiados por las cascadas de partículas que generan en la atmósfera. Dado que la mayoría de las cascadas son iniciadas por hadrones con energías que van desde 1000 GeV hasta energías mayores a 10 TeV , que al entrar isotrópicamente a la atmósfera producen un gran número de productos secundarios en una serie de colisiones sucesivas con los núcleos de los constituyentes atmosféricos ie. N_2 , O_2 , Ar (Figura 2.1). Estos productos secundarios resultantes se comportan de una manera similar, mientras se van propagando a través de la atmósfera.

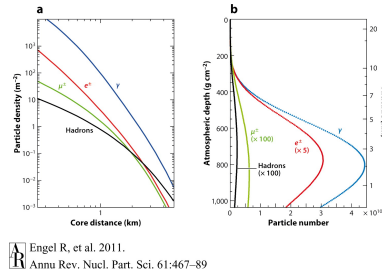


FIGURA 2.2: Distribución longitudinal promedio de las cascadas inducidas por un protón a 10^{19} GeV. Figura obtenida de [4].

Las cascadas de partículas descritas anteriormente son conocidas como **cascadas atmosféricas extensas EAS** [10]. Estas cascadas atmosféricas se propagan longitudinalmente a lo largo de la dirección de incidencia de la partícula primaria, debido al momento transversal de los productos secundarios la cascada también se extiende lateralmente. Como referencia de la complejidad que se tiene, una partícula primaria altamente energética puede crear una cascada gigante de partículas que se propaga esencialmente a la velocidad de la luz a través de la atmósfera y puede alcanzar el nivel del mar si el evento es lo suficientemente energético, del mismo modo, el número de productos secundarios es del orden $O(10^{10})$ (Figura 2.2).

2.2.1. Propiedades de las cascadas atmosféricas

Algunas propiedades que caracterizan a las cascadas atmosféricas son:

- E_0 [eV]: energía de la partícula primaria.
- N (*shower size*): número total de partículas producidas en un nivel en particular de la atmósfera. Es una función que depende de la energía E_0 el ángulo de incidencia cenit y a altura de la primera interacción del evento primario en la atmósfera h_1 .
- X_{\max} [gcm^{-2}]: profundidad de máximo desarrollo, medida desde la parte más alta de la atmósfera. Se desfasa a profundidades mayores conforme la energía del primario se incrementa.
- **Shower Axis**: extensión del vector de momento del primario incidente en la dirección de propagación de la cascada.
- **Dirección de arribo**: dirección de incidencia de la partícula primaria determinada por sus ángulos acimut y cenit.

2.2.2. Métodos de detección

Los principales detectores de cascadas atmosféricas son arreglos de detectores espaciados uno del otro en distancias que dependen de la energía de los rayos cósmicos que se quieren observar. Para energías de 10^6 GeV, la distancia entre los detectores deben de ser del orden de decenas de metros y para energías que sobrepasan 10^9 GeV, la distancia es del orden de miles de metros, por ejemplo en el observatorio Pierre Auger la distancia entre detectores es de 1500 m. Diferentes métodos observacionales como detectores de aire cherenkov o detectores de fluorescencia son combinados con estos arreglos como en el caso del observatorio Pierre Auger (Figura 2.3).

Sin importar el tipo de sistema de detección que se use, los datos adquiridos representan a la cascada en una etapa en particular de su desarrollo, así como una foto instantánea de la

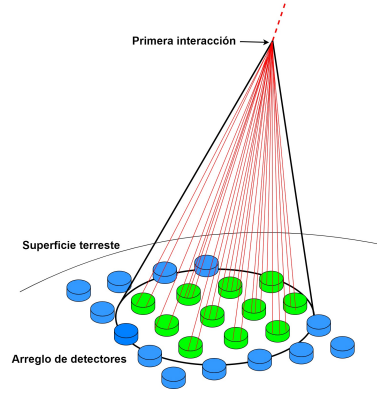


FIGURA 2.3: Esquema de una cascada atmosférica extensa. El rayo cósmico primario (línea punteada) interactúa con constituyentes atmosféricos (a una altura típica de ~ 20 km) y produce una cascada de partículas elementales (líneas rojas sólidas). Un arreglo de detectores (que cubren áreas alrededor $\sim km^2$) detectan el arribo de las partículas.

cascada, en el plano de observación. Con estos datos se puede conocer información básica que caracteriza a la cascada atmosférica, así como los tiempos de arribo de las partículas cargadas, fotones asociados no ópticos, las distribuciones laterales de partículas y fotones en el plano de observación a una profundidad atmosférica específica. Las propiedades anteriores son inmediatamente accesibles con un arreglo de detectores simple.

La reconstrucción de la partícula primaria depende del modelo hadrónico de interacciones que use el simulador Monte Carlo.

2.3. Análisis de las cascadas atmosféricas

Las cascadas atmosféricas están caracterizadas por un delgado disco de partículas radialmente extenso que se propaga a la velocidad de la luz a través del eje de la cascada. El patrón de las cascadas es circular para cascadas verticalmente incidentes, mientras que la extensión longitudinal y lateral dependen principalmente de la energía de la partícula primaria. El esparcimiento lateral de las partículas en regiones bajas de la atmósfera, llega a cubrir áreas de varios kilómetros cuadrados, agregado a lo anterior, la mayoría de las partículas arriban en intervalos estrechos de tiempo que van desde unos cuantos nanosegundos, en la vecindad del eje de la cascada, hasta unos 10 ns a distancias mayores del núcleo de la cascada.

Eventos de baja energía alcanzan su máximo desarrollo en zonas altas de la atmósfera y se mitigan lentamente a mayor profundidad; los componentes que alcanzan a llegar a la superficie son los muones y neutrinos. Para eventos extremadamente energéticos las cascadas logran alcanzar su máximo desarrollo a nivel del mar mientras que sus componentes hadrónicos y electromagnéticos sobrevivientes, son absorbidos en la superficie terrestre y los muones resultantes altamente energéticos continúan propagándose bajo tierra.

Como regla de dedo se puede decir que en promedio una cascada atmosférica está constituida al 1 % por hadrones, alrededor del 10 % son muones y el 90 % o más son electrones o positrones. También para las primeras estimaciones energéticas de la primaria, se tiene que cascadas verticales a una altitud de 5 km tienen una energía 1 GeV, 3 GeV para alturas entre $2,5$ km y 3 km 10 GeV a nivel del mar.

Para poder visualizar la ocurrencia de estos eventos en distintos regímenes energéticos se puede estudiar el número de partículas que cruzan un área en un tiempo dado, conocido como flujo cósmico. Este flujo sigue una ley de potencias con la forma $\frac{1}{E^3}$, en Figura 2.4 se

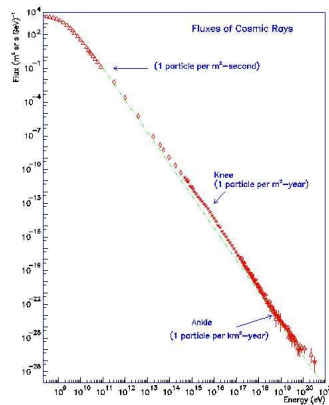


FIGURA 2.4: Espectro de energía de los rayos cósmicos.
Figura obtenida de [13].

puede ver energías alrededor de 10^{12} eV el flujo es de 10 partículas primarias por minuto y m^2 , para energías entre 10^{18} eV y 10^{19} eV se tiene 1 partícula primaria por año y km^2 , en estos regímenes energéticos las estadísticas son pobres y las incertidumbres altas.

Los observables importantes que se deben de obtener para poder reconstruir una cascada son, el tiempo de arribo t_i al detector i respecto al tiempo de referencia t_0 , la densidad de partículas ρ_i y la posición del detector con respecto al sistema de referencia (x_i, y_i) . Con la distribución lateral de las partículas su puede adquirir la localización del eje de la cascada y el tamaño de la cascada para así obtener un estimado de la energía. Los parámetros anteriores son clasificados como accesibles directamente ya que no necesitan un análisis complejo para su adquisición.

Por último, parámetros indirectamente accesibles que se relacionan con la naturaleza de la partícula primaria, como el tipo de partícula, masa y carga, no se pueden extraer inmediatamente y requieren métodos sofisticados de análisis.

2.4. Simulaciones

Los simuladores de cascadas atmosféricas son de vital importancia para la evaluación e interpretación de datos experimentales. Las técnicas se reducen a crear e insertar un modelo de cascadas que corresponda a nuestro mejor entendimiento de la realidad, simular cascadas, comparar resultados con los datos experimentales, modificar el modelo o sus parámetros y intentar de nuevo; hacer ajustes pequeños al modelo y repetir hasta que se obtenga un consenso entre la predicción y el experimento.

Las cascadas iniciadas por primarios hadrónicos consisten en la superposición de dos tipos de cascadas, una hadrónica y otra electromagnética. La cascada electromagnética se entiende bien y solo posee problemas prácticos asociados al gran número de partículas participantes en el orden de $\sim 10^{10}$. Los programas computacionales que simulan cascadas hadrónicas o electromagnéticas de alta energía o cascadas atmosféricas completas son altamente complejos.

Para tomar en cuenta la complejidad computacional que se tiene, una simulación completa de una cascada debe incluir los componentes electromagnéticos y hadrónicos. Además se debe de tomar en cuenta todos los procesos relevantes, donde la mayoría son de naturaleza estocástica y muchos están en competencia entre sí. También se debe incluir los parámetros que especifican el estado de cada partícula como su masa, carga, energía, momento,

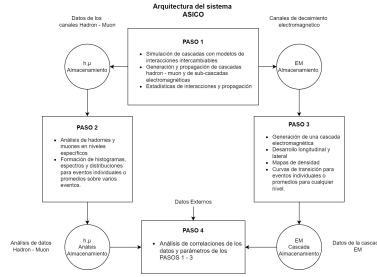


FIGURA 2.5: Diagrama de la arquitectura ASICO.

ubicación de su creación en el espacio tiempo x, y, z, t , la orientación angular respecto al marco de referencia y parámetros genéticos que revelan la altura de la interacción donde cada partícula fue creada. Estos observables son cruciales para análisis subsecuentes y para la comparación con datos experimentales.

2.4.1. Estrategia para simulaciones de EAS

La arquitectura del sistema ASICO sirve como base para entender el proceso de un simulador completo. ASICO fue el primer simulador que generaba cascadas detalladas al usar 12 parámetros que definen a cada partícula y este fue la base para el desarrollo de CORSIKA. Para simular una cascada completa se comienza con la simulación de la cascada hadrónica y se le llama *PASO 1*, este paso da lugar a datos como la elasticidad, distribuciones de interacciones hadrónicas en diferentes rangos de energías, entre otros; para la parte electromagnética de la cascada se le llama *PASO 2* y la combinación de las dos simulaciones para formar la cascada completa se le conoce como el *PASO 3* [10].

2.4.2. Problemática de las simulaciones

Los parámetros que determinan a cada partícula se deben de asignar en su punto de creación y requieren actualizarse después de cada proceso al cual está sujeta. Esto es al final de cada trayectoria particular, después de propagarse al siguiente punto de interacción o decaimiento y cuando pasan a un nuevo nivel de observación. En cada actualización la partícula y sus parámetros son guardados para su subsecuente análisis y evaluación de los datos de la cascada simulada.

Consecuentemente, la ejecución de programas que simulan cascadas requieren mucho tiempo computacional particularmente para cascadas energéticas donde el número de partículas involucradas se vuelve muy grande. La gran cantidad de datos producidos por estas simulaciones requieren también una gran capacidad de almacenamiento. La complejidad aumenta si los componentes atmosféricos cherenkov o de fluorescencia se incluyen, en este caso los datos corren el riesgo de divergir y métodos computacionales más sofisticados deben de ser usados para su análisis.

En general el desarrollo y la propagación de una cascada a partir del punto de iniciación (la primera interacción) hasta el nivel de observación consume más tiempo que el análisis subsecuente de los datos producidos.

- **Memoria:** Almacenamiento confiable de los datos crudos.
- **Rastreabilidad:** Rastreo de los parámetros que determinan a cada partícula.
- **Fenomenología:** Propagación de las partículas en la atmósfera tomando en cuenta todos los procesos a los que están sujetas.

- **Configuración inicial:** Correlación confiable entre los parámetros iniciales (*i.e.* modelos de interacción, propiedades del detector, propiedades de la primaria) con la cascada final
- **Tiempo:** Existe una relación lineal entre la energía de la partícula y el tiempo que lleva simular la cascada que genera.

2.5. Modelos generativos

Los modelos generativos son un tipo de aprendizaje no supervisado, que describen cómo se genera un conjunto de datos en términos de un modelo probabilístico. Al muestrear de dicho modelo se es capaz de generar datos no observados previamente. Típicamente el marco de trabajo de los modelos generativos involucra las siguientes partes [8].

- **Los datos:** Conjunto de observaciones, que se asumen ser generadas de acuerdo a una distribución de probabilidad desconocida.
- **El modelo:** Un modelo generativo que intenta imitar lo mejor posible, a la distribución que genera las observaciones. Este modelo es capaz de generar datos no observados que parecen haber sido generados con la distribución desconocida y no debe de generar datos conocidos.

2.5.1. Aplicación de modelos generativos

Algunas de las tareas modernas de los modelos generativos son:

- **Generación de datos novedosos:** Se generan datos nunca antes vistos que pueden ser utilizados para imitar fenómenos o para ayudar a flujos en modelos discriminativos con un pre entrenamiento autosupervisado.
- **Compresión de datos:** El modelo es capaz de aprender las características más importantes que determinan a la observación y así logra reducir la dimensionalidad del espacio de características donde vive la observación original.
- **Tecnologías de síntesis condicional:** Proporciona un método capaz de generar información novedosa condicionada a un dominio específico. Lo anterior permite una suerte de transformación de datos de un dominio a otro.

Las tareas anteriores han logrado avances en:

- Generación de rostros humanos
- Transformación de imágenes
- Transferencia de estilos
- Texto a imagen
- Texto a voz
- Edición de imágenes
- Super resolución
- Generación de objetos 3D
- Predicción de fotogramas en videos

Capítulo 3

Modelos generativos

Este capítulo proporciona una breve y concreta introducción a los modelos generativos, su fundamento teórico y algunas de las principales áreas de investigación. Además de mostrar las principales aplicaciones en HEP.

3.1. Introducción

Así como los modelos discriminativos han sido el centro del progreso en metodologías del aprendizaje máquina en los últimos años ya que es más sencillo monitorear su desempeño y así poder elegir la mejor metodología que se ajuste a la tarea. Los modelos generativos suelen ser más difíciles de evaluar, lo cual hace que encontrar aplicaciones industriales sea más complicado.

Los modelos generativos han probado su efectividad para generar muestras que son capaces de imitar a observaciones reales así como rostros humanos con StyleGAN de NVIDIA o GPT3 de openAI para generar texto. Los modelos anteriores han impulsado el interés para expandir el campo del aprendizaje de máquina a través de modelos que aprenden a generar muestras indistinguibles de observaciones reales. Los avances en el campo podrían ser fundamentales para el desarrollo de una máquina que haya adquirido una inteligencia comparable a la de los humanos.

El campo de los modelos generativos es diverso y la definición de los problemas toman una gran variedad de formas. Sin embargo, para cada tarea, los desafíos que se tienen son los mismos. Entender cómo es que el modelo maneja un alto grado de dependencias condicionales entre las características de la observación y como es que logra encontrar una observación viable, en un espacio de alta dimensionalidad, que concuerda con los datos observados a partir de un conjunto pequeño de observaciones, es de vital importancia para desarrollar metodologías más robustas.

3.1.1. Fundamento teórico

Como punto de partida se debe de reconocer la diferencia clave entre los modelos discriminativos y los modelos generativos.

Los modelos discriminativos estiman $p(y|x)$ - la probabilidad del observable y dada la observación x .

Los modelos generativos estiman $p(x)$ - la probabilidad de observar x .

En otras palabras, los modelos discriminativos intentan estimar la probabilidad de que una observación x pertenezca a la categoría y , y los modelos generativos intentan estimar la probabilidad de ver la observación x . Por lo tanto el modelo debe de ser probabilístico en vez de ser determinista y debe incluir un elemento estocástico que influencia las observaciones individuales generadas por el modelo.

3.2. Líneas de investigación

Detrás del creciente interés en la academia por los modelos generativos se encuentran dos razones con una gran importancia teórica, que se describen a continuación:

Se debe de buscar un entendimiento completo de cómo se generan las observaciones para así poder formar inteligencias artificiales más sofisticadas que van más allá de lo que pueden lograr los modelos discriminativos.

Es altamente probable que los modelos generativos sean centrales para futuros desarrollos en otros campos del aprendizaje máquina.

3.3. Modelos generativos

Un modelo generativo describe cómo se genera un conjunto de observaciones en términos de un modelo probabilístico. Al generar muestras de este modelo, se es capaz de generar observaciones nunca antes vistas.

Al tener un conjunto de observaciones que representen la entidad que se quiere generar. El objetivo del modelo generativo es generar nuevas observaciones que sigan las mismas reglas con las cuales las observaciones originales fueron generadas. Lo anterior es posible debido a que se asume que existe alguna distribución de probabilidad que explica, porqué ciertas observaciones son más probables de encontrarse en un conjunto y otras no.

El trabajo del modelo es imitar una distribución desconocida lo más cercano posible, para luego muestrear de ella y generar nuevas observaciones, distintas de las conocidas, que además parezca que son parte del conjunto de entrenamiento.

El marco de trabajo de los modelos generativos es el siguiente:

- Se tiene un conjunto de datos de observaciones X .
- Se asume que las observaciones X se generaron de acuerdo a una distribución de probabilidad desconocida p_{datos} .
- Se diseña un modelo generativo p_{modelo} que intenta imitar a p_{datos} .
- Se muestrea de p_{modelo} para generar observaciones que parecen ser obtenidas de p_{datos} .

Consideramos que p_{modelo} hace un buen trabajo si:

Puede generar observaciones que parecen ser obtenidas de p_{datos} .

Puede generar observaciones que son sustancialmente diferentes a las observadas en X . En otras palabras, el modelo no debería de reproducir cosas que ya conoce.

3.3.1. Conceptos básicos

Espacio Muestral

El espacio muestral es el conjunto de todos los posibles valores que una observación x puede tomar.

Función de densidad de probabilidad

Una función de densidad de probabilidad, $p(x)$, es una función que mapea un punto x del espacio muestral a un número entre 0 y 1. La suma de la función de densidad sobre todos los puntos del espacio muestral es igual a 1 para que sea una distribución bien definida. Por definición tenemos que solo existe una p_{datos} pero existen infinitas distribuciones p_{modelo} que pueden estimar p_{datos} . Para encontrar una distribución adecuada se tiene que usar un modelo paramétrico.

Modelo paramétrico

Un modelo paramétrico, $p_{\theta}(x)$, es una familia de funciones de densidad que pueden ser descritas por medio de un número finito de parámetros, θ .

Verosimilitud

La verosimilitud $L(\theta|x)$ de un conjunto de parámetros θ es una función que mide la plausibilidad de θ , dado una observación x .

Se define como sigue:

$$L(\theta|x) = p_{\theta}(x)$$

Esto es, la verosimilitud de θ dada una observación x está definida como el valor de la función de densidad de probabilidad parametrizada por θ , en el punto x .

Para un conjunto de observaciones se tiene:

$$L(\theta|X) = \prod_{x \in X} p_{\theta}(x)$$

La expresión anterior resulta ser complicada de trabajar computacionalmente hablando, entonces se usa la verosimilitud logarítmica l .

$$l(\theta|x) = \sum_{x \in X} \log p_{\theta}(x)$$

En otras palabras, la verosimilitud de un conjunto de parámetros θ es igual a la probabilidad de observar los datos bajo el modelo parametrizado por θ .

3.3.2. Modelos basados en la verosimilitud

Dado un conjunto de observaciones finito D , muestreado de una distribución p_{datos} , el objetivo de un modelo generativo es aproximar p_{datos} a través de una aproximación paramétrica p_{θ} , de tal manera que, se puede pensar que el modelo aprende los parámetros que minimicen alguna suerte de distancia entre p_{datos} y p_{θ} (). Matemáticamente se tiene el siguiente problema de optimización.

$$\min_{\theta \in \mathcal{M}} d(p_{datos}, p_{\theta}) \quad (3.1)$$

La verosimilitud es la primera métrica que se considera para medir la distancia entre el modelo y p_{datos} . A esta técnica se le conoce como estimación de la verosimilitud máxima (MLE).

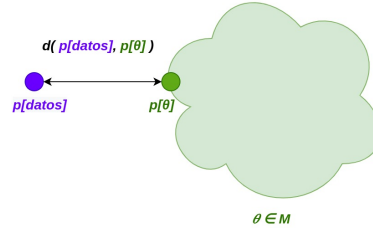


FIGURA 3.1: Caption

MLE

Esta técnica permite estimar el conjunto de parámetros $\hat{\theta}$ del modelo $p_{\theta}(x)$, los cuales son los más probables de explicar los datos observados X .

Formalmente:

$$\begin{aligned}\hat{\theta} &= \arg \max_{\theta} l(\theta|X) \\ &= \arg \max_{\theta} \sum_{x \in X} \log p_{\theta}(x)\end{aligned}\tag{3.2}$$

lo cual es equivalente a minimizar la divergencia KL.

$$\hat{\theta} = \arg \min_{\theta} D_{KL}(P_{datos} || P_{\theta})\tag{3.3}$$

Para los modelos basados en la verosimilitud, se busca que se ajusten lo mejor que se pueda a los datos de entrenamiento o que idealmente, logren representar idénticamente a la distribución real p_{datos} y para nuevas observaciones x , el modelo debe de ser capaz de evaluar $p_{\theta}(x)$ de igual forma se debe de poder muestrear de la distribución $p_{\theta}(x)$. Por último el modelo debe obtener una representación latente significativa.

Los objetivos principales de los modelos basados en la verosimilitud son:

- Muestreo rápido
- Inferencia rápida
- Entrenamiento rápido
- Buena calidad de las muestras
- Buena compresión

Algunas de las clases más grandes de estos modelos

- Modelos de flujo
 - Flujos autorregresivos rápido para evaluar x arbitrarias, lento para muestrear
 - Flujos inversos autorregresivos lento para evaluar x arbitrarias, rápido para muestrear
- Modelos de variable latente
 - VAE lento para muestrear

A grandes rasgos los pasos que estos modelos siguen, son muestrear, evaluar la verosimilitud, entrenar y obtener una representación latente, pero si solo importara obtener muestras confiables se debe introducir la clase de modelos generativos implícitos.

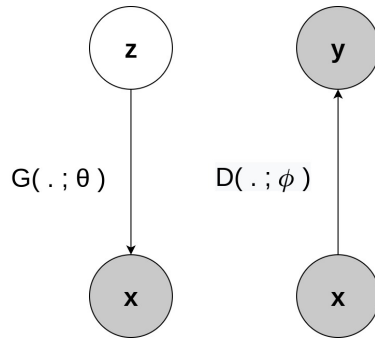


FIGURA 3.2: caption

3.3.3. Modelos implícitos

Los modelos implícitos no estiman una densidad de probabilidad explícitamente, solo se enfocan en la calidad de muestras generadas. De acuerdo al marco de trabajo de los modelos generativos, se comienza por tener muestras de una distribución $p_{\text{datos}} : x_1, x_2, \dots, x_n$ luego se define un modelo paramétrico $q_\phi(z) = DNN(z\phi)$ que muestrea a partir de un vector latente muestreado de una fuente fijo de ruido (uniforme o gaussiana) $z \sim p(z)$.

Lo anterior quiere decir que $q_\phi(z)$ induce un modelo de densidad p_{modelo} , sin embargo, no se tiene una forma explícita de p_{datos} o p_{modelo} , solo se tiene acceso a las observaciones que generan. Lo anterior obliga a buscar una medida de distancia entre p_{datos} y p_{modelo} que no necesite la expresión explícita de p_{modelo} y que se comporte diferente a la estimación de verosimilitud máxima. Algunos ejemplos de dichas métricas son la divergencia de Jensen Shannon JSD o la métrica Earth Mover's Distance EMD.

GANs

Los modelos GAN son el primer ejemplo de un modelo que no se basa en MLE por consiguiente se estima el modelo generativo mediante un proceso adversario, el cual debe entrenar dos modelos simultáneamente

- Un modelo generativo de variable latente G_θ que captura la distribución de los datos y genera muestras x a partir de z determinísticamente.
- Un modelo discriminativo D_ϕ que estima la probabilidad de que una haya sido generada por p_{datos} o por G_θ .

Formalmente el objetivo de la GAN se escribe como sigue.

$$\min_{\theta} \max_{\phi} \mathbb{E}_{x \sim p_{\text{datos}}} [\log D_\phi(x)] + \mathbb{E}_{z \sim p(z)} [\log(1 - D_\phi(G_\theta(z)))] \quad (3.4)$$

La función objetivo describe un juego minimax entre el generador G y el discriminador D , donde G intenta minimizar la probabilidad de que sus muestras sean clasificadas como falsas por D ($p_{\text{datos}} = p_\theta$) y D intenta maximizar la probabilidad para el problema de clasificación binaria ($p_{\text{data}} \neq p_\theta$).

La noción de distancia que se tiene con el objetivo está relacionada con JSD al considerar al discriminador D óptimo se tiene que el generador intenta optimizar la siguiente expresión.

$$V(G, D^*) = -\log 4 + 2D_{JSD}(p_{\text{datos}} || p_G) \quad (3.5)$$

Nos dice que para esta métrica, el generador óptimo para la función objetivo de la GAN se vuelve $p_G = p_{\text{datos}}$. Por último, para evitar la saturación de D se reformula las funciones de pérdida como sigue:

$$L^{(D)} = -\mathbb{E}_{x \sim p_{\text{datos}}} [\log D_\phi(x)] + \mathbb{E}_{z \sim p(z)} [\log(1 - D_\phi(G_\theta(z)))] \quad (3.6)$$

$$L^{(G)} = -\mathbb{E}_{z \sim p(z)} [\log D_\phi(G_\theta(z))] \quad (3.7)$$

La evaluación de estos modelos aún es un problema abierto, a diferencia de otros modelos no se puede reportar estimados de la verosimilitud explícitos en conjuntos de evaluación, sin embargo, estos modelos son rápidos para muestrear, generan muestras sorprendentemente similares y interpolan correctamente entre las muestras.

WGAN

La arquitectura WGAN tiene como objetivo minimizar EMD y tiene la bondad de mejorar la calidad de las muestras generadas. Su correspondiente función objetivo es.

$$\min_G \max_{D \in \mathcal{D}} \mathbb{E}_{x \sim p_{\text{datos}}} [D(x)] + \mathbb{E}_{\tilde{x} \sim P_G} [D(\tilde{x})] \quad (3.8)$$

Este nuevo objetivo usa una nueva métrica para optimizar el generador que ataca las inestabilidades que surgen al usar JSD, de igual manera introduce la noción de continuidad de Lipschitz para estabilizar el entrenamiento de la GAN. En su versión donde la continuidad de Lipschitz se asegura mediante un término regularizador el objetivo cambia a

$$\min_G \max_{D \in \mathcal{D}} \mathbb{E}_{x \sim p_{\text{datos}}} [D(x)] + \mathbb{E}_{\tilde{x} \sim P_G} [D(\tilde{x})] + \lambda \mathbb{E}_{\hat{x} \sim P_{\hat{x}}} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \quad (3.9)$$

donde:

$$\hat{x} \leftarrow \epsilon x + (1 - \epsilon)\tilde{x}$$

Conditional GAN

$$\min_G \max_D \mathbb{E}_{x,y} [\log D(x, G(x)) + \log(1 - D(x, y))] \quad (3.10)$$

3.3.4. Aprendizaje de representación

La idea principal detrás de aprendizaje representation es que en vez de tratar de modelar directamente el espacio muestral en altas dimensiones, se debe de intentar describir cada observación en el conjunto de entrenamiento, usando un espacio latente de menor dimensión para luego aprender una función que mapea puntos en el espacio latente al dominio original.

Hay que notar que el aprendizaje de representación no solo asigna valores a un conjunto de características del espacio latente para alguna observación si no que, aprende qué características son más importantes para poder describir las observaciones dadas y cómo generar estas características a partir del conjunto de entrenamiento.

Algunos de los modelos que se especializan en son la InfoGAN y la BiGAN.

Capítulo 4

Método experimental

En este capítulo se detallan las técnicas utilizadas así como el método experimental para la generación y evaluación de cascadas atmosféricas..

4.1. Recolección y manejo de datos

Dado que el objetivo de este trabajo es implementar un modelo generativo que sea capaz de generar simulaciones de cascadas atmosféricas, se necesita un banco de datos de dichas simulaciones. Una de las primeras maneras de obtener este conjunto de datos será tomar el simulador CORSIKA y generar el número de eventos necesarios para poder entrenar a la red neuronal. Como segunda opción de recolección se buscará un banco de simulaciones disponibles al público con características fundamentales como la energía de la primaria, la altura de máximo desarrollo, etc.

Como se ha mencionado en capítulos anteriores, las tareas que cada simulador conlleva para generar una cascada, incluyen subrutinas que se encargan de la parte hadrónica, la parte electromagnética, el modelo de interacción, el rastreo de parámetros y procesamientos adicionales para reducir la carga computacional; por lo cual los datos que se utilizarán no deben concentrarse en alguna de estas subrutinas para poder acotar el dominio de la simulación.

Para generar un resultado significativo sin pérdida de generalidad, se tomarán en cuenta las siguientes consideraciones:

Utilizar un número relativamente “grande” de simulaciones. El número de observaciones debe ser un número “grande”, como referencia se intentará obtener un conjunto de datos con al menos el mismo número de observaciones que el conjunto MNIST.

Preprocesar el conjunto de datos. Para datos como los tiempos de arribo, el dominio de cada simulación diferirá en magnitudes que deben tomarse en cuenta, por lo tanto técnicas de estandarización podrán ser utilizadas para tener una mejor estabilidad al entrenar el modelo.

Separación en conjuntos de entrenamiento, validación y evaluación. Una vez que se cumpla lo anterior, los datos estarán listos para poder entrenar al modelo.

Algorithm 1 Algoritmo de entrenamiento de la GAN

```

Require:  $k \geq 1$ 
for  $N$  iteraciones do
  end for
 $y \leftarrow 1$ 
 $X \leftarrow x$ 
 $N \leftarrow n$ 
while  $N \neq 0$  do
  if  $N$  is even then
     $X \leftarrow X \times X$ 
     $N \leftarrow \frac{N}{2}$ 
  else if  $N$  is odd then
     $y \leftarrow y \times X$ 
     $N \leftarrow N - 1$ 
  end if
end while

```

▷ This is a comment

4.2. Algoritmos

4.3. Desarrollo del método

A continuación se muestra el método desglosado. Obtener un banco de datos de simulaciones de cascadas atmosféricas acotadas a algún proceso seleccionado.

Hacer el preprocesamiento de datos adecuado para normalizar la entrada de datos y así estabilizar el entrenamiento sin dejar de tomar en cuenta las consideraciones teóricas que dicho procesamiento implica.

Elegir una arquitectura generativa adversaria y entrenarla para intentar generar nuevas cascadas.

Se tomarán las nuevas muestras y se inspeccionarán visualmente comparándolas con las observaciones originales.

Después de haber confirmado la similitud con las observaciones originales, se hará un análisis exploratorio de la dimensionalidad del vector latente y un análisis al proceso determinista de entradas y salidas de dicha red.

Una vez logrado lo anterior, cambiar de arquitectura y repetir el proceso, si se puede.

4.3.1. Mapa del método experimental

Capítulo 5

Resultados

5.1. Dsicusion de resultados obtenidos

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

Capítulo 6

Conlusiones

6.1. Conclusiones puntuales obtenidas

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

6.2. Aplicacion y extension generadas del trabajo

6.3. Trabajos futuros

Capítulo 7

LaTeX Review

7.1. Welcome and Thank You

Welcome to this \LaTeX Thesis Template, a beautiful and easy to use template for writing a thesis using the \LaTeX typesetting system.

If you are writing a thesis (or will be in the future) and its subject is technical or mathematical (though it doesn't have to be), then creating it in \LaTeX is highly recommended as a way to make sure you can just get down to the essential writing without having to worry over formatting or wasting time arguing with your word processor.

\LaTeX is easily able to professionally typeset documents that run to hundreds or thousands of pages long. With simple mark-up commands, it automatically sets out the table of contents, margins, page headers and footers and keeps the formatting consistent and beautiful. One of its main strengths is the way it can easily typeset mathematics, even *heavy* mathematics. Even if those equations are the most horribly twisted and most difficult mathematical problems that can only be solved on a super-computer, you can at least count on \LaTeX to make them look stunning.

7.2. Learning \LaTeX

\LaTeX is not a WYSIWYG (What You See is What You Get) program, unlike word processors such as Microsoft Word or Apple's Pages. Instead, a document written for \LaTeX is actually a simple, plain text file that contains *no formatting*. You tell \LaTeX how you want the formatting in the finished document by writing in simple commands amongst the text, for example, if I want to use *italic text for emphasis*, I write the `\emph{text}` command and put the text I want in italics in between the curly braces. This means that \LaTeX is a «mark-up» language, very much like HTML.

```

1 from sys import argv
2
3 def reader(file):
4     """
5     This function returns a matrix M[i][j] of the input file.csv.
6     """
7     lines = open(file, 'r').readlines()
8     a = []
9     for i in range(0, len(lines)):
10         a.append(lines[i].strip().split(','))
11     return a
12
13 def reader_grammar(file):
14     lines = open(file, 'r').readlines()
15     a = []
16     for i in range(0, len(lines)):

```

```

17     a.append(lines[i].strip().split('>'))
18     return a
19
20 def states(matr_of_file):
21     """
22     This function returns a list of the states of a given transition matrix of a
23     machine.
24     The input must be a list of list of the form M[i][j].
25     """
26     a = []
27     for i in range(1, len(matr_of_file)):
28         a.append(matr_of_file[i][0])
29     return a
30
31 def alphabet(matr_of_file):
32     """
33     This function returns a list of the alphabet of the machine M given by the
34     input in matrix form M[i][j].
35     """
36     a = []
37     for i in range(1, len(matr_of_file[0])-1):
38         a.append(matr_of_file[0][i])
39     return a
40
41 def number_list(l):
42     """
43     This function returns a list of number from 1 to len(l) given the list l.
44     """
45     a = []
46     for x in range(1, len(l)+1):
47         a.append(x)
48     return a
49
50 def trans_func(matr_of_file, state, char):
51     """
52     This function returns the state that the machine goes to, given an state and
53     a character of the alphabet.
54     e.g. trans_funciton('state[0]', 'a') = 'state1'
55     """
56     return matr_of_file[dict(zip(states(matr_of_file), number_list(states(
57         matr_of_file))))[state]][dict(zip(alphabet(matr_of_file), number_list(
58         alphabet(matr_of_file))))[char]]
59
60 def init_state(matr_of_file):
61     """
62     This function returns the initial state of the machine M given by the input in
63     matrix form M[i][j].
64     """
65     return matr_of_file[1][0]
66
67 def eof(matr_of_file):
68     """
69     This function returns a list of the acceptance states of the machine M given
70     by the input in matrix form M[i][j].
71     """
72     a = []
73     for i in range(0, len(states(matr_of_file))):
74         if matr_of_file[i+1][len(matr_of_file[0])-1] == 'accept':
75             a.append(states(matr_of_file)[i])
76     return a
77
78 def isinalpha(a, alph):
79     """
80     This is a boolean function that tells if the character a is in the alphabet
81     alph.
82     e.g. isinalpha('a', ['a', 'b']) = True
83     """
84     return a in alph
85
86 def machine(file, string):
87     """
88     This function implements the SFA given the input file in matrix form M[i][j]
89     and given a string.

```



```

81 e.g. machine(reader('filename.csv'), 'hola') = No
82 """
83 if isinalpha(string[0], alphabet(file)):
84     s0 = trans_func(file, init_state(file), string[0])
85     i = 0
86     while i < len(string)-1:
87         n = s0
88         i += 1
89         if isinalpha(string[i], alphabet(file)):
90             try:
91                 s0 = trans_func(file, n, string[i])
92             except:
93                 return 'no'
94         else:
95             return 'no'
96     if (s0 in eof(file)) == True:
97         return 'yes'
98     else:
99         return 'no'
100 else:
101     return 'no'
102
103 script, file, string = argv
104 print(machine(reader(file), string))

```

7.2.1. A (not so short) Introduction to L^AT_EX

If you are new to L^AT_EX, there is a very good eBook – freely available online as a PDF file – called, «The Not So Short Introduction to L^AT_EX». The book's title is typically shortened to just *lshort*. You can download the latest version (as it is occasionally updated) from here: <http://www.ctan.org/tex-archive/info/lshort/english/lshort.pdf>

It is also available in several other languages. Find yours from the list on this page: <http://www.ctan.org/tex-archive/info/lshort/>

It is recommended to take a little time out to learn how to use L^AT_EX by creating several, small 'test' documents, or having a close look at several templates on:

<http://www.LaTeXTemplates.com>

Making the effort now means you're not stuck learning the system when what you *really* need to be doing is writing your thesis.

7.2.2. A Short Math Guide for L^AT_EX

If you are writing a technical or mathematical thesis, then you may want to read the document by the AMS (American Mathematical Society) called, «A Short Math Guide for L^AT_EX». It can be found online here: <http://www.ams.org/tex/amslatex.html> under the «Additional Documentation» section towards the bottom of the page.

7.2.3. Common L^AT_EX Math Symbols

There are a multitude of mathematical symbols available for L^AT_EX and it would take a great effort to learn the commands for them all. The most common ones you are likely to use are shown on this page: <http://www.sunilpatel.co.uk/latex-type/latex-math-symbols/>

You can use this page as a reference or crib sheet, the symbols are rendered as large, high quality images so you can quickly find the L^AT_EX command for the symbol you need.

7.2.4. L^AT_EX on a Mac

The L^AT_EX distribution is available for many systems including Windows, Linux and Mac OS X. The package for OS X is called MacTeX and it contains all the applications you need – bundled together and pre-customized – for a fully working L^AT_EX environment and work flow.

MacTeX includes a custom dedicated L^AT_EX editor called TeXShop for writing your ‘.tex’ files and BibDesk: a program to manage your references and create your bibliography section just as easily as managing songs and creating playlists in iTunes.

7.3. Getting Started with this Template

If you are familiar with L^AT_EX, then you should explore the directory structure of the template and then proceed to place your own information into the *THESIS INFORMATION* block of the `main.tex` file. You can then modify the rest of this file to your unique specifications based on your degree/university. Section 7.5 on page 32 will help you do this. Make sure you also read section 7.7 about thesis conventions to get the most out of this template.

If you are new to L^AT_EX it is recommended that you carry on reading through the rest of the information in this document.

Before you begin using this template you should ensure that its style complies with the thesis style guidelines imposed by your institution. In most cases this template style and layout will be suitable. If it is not, it may only require a small change to bring the template in line with your institution’s recommendations. These modifications will need to be done on the `MastersDoctoralThesis.cls` file.

7.3.1. About this Template

This L^AT_EX Thesis Template is originally based and created around a L^AT_EX style file created by Steve R. Gunn from the University of Southampton (UK), department of Electronics and Computer Science. You can find his original thesis style file at his site, here: <http://www.ecs.soton.ac.uk/~srg/softwaretools/document/templates/>

Steve’s `ecsthesis.cls` was then taken by Sunil Patel who modified it by creating a skeleton framework and folder structure to place the thesis files in. The resulting template can be found on Sunil’s site here: <http://www.sunilpatel.co.uk/thesis-template>

Sunil’s template was made available through <http://www.LaTeXTemplates.com> where it was modified many times based on user requests and questions. Version 2.0 and onwards of this template represents a major modification to Sunil’s template and is, in fact, hardly recognisable. The work to make version 2.0 possible was carried out by Vel and Johannes Böttcher.

7.4. What this Template Includes

7.4.1. Folders

This template comes as a single zip file that expands out to several files and folders. The folder names are mostly self-explanatory:

Appendices – this is the folder where you put the appendices. Each appendix should go into its own separate `.tex` file. An example and template are included in the directory.

Chapters – this is the folder where you put the thesis chapters. A thesis usually has about six chapters, though there is no hard rule on this. Each chapter should go in its own separate `.tex` file and they can be split as:

- Chapter 1: Introduction to the thesis topic
- Chapter 2: Background information and theory
- Chapter 3: (Laboratory) experimental setup
- Chapter 4: Details of experiment 1
- Chapter 5: Details of experiment 2
- Chapter 6: Discussion of the experimental results
- Chapter 7: Conclusion and future directions

This chapter layout is specialised for the experimental sciences, your discipline may be different.

Figures – this folder contains all figures for the thesis. These are the final images that will go into the thesis document.

7.4.2. Files

Included are also several files, most of them are plain text and you can see their contents in a text editor. After initial compilation, you will see that more auxiliary files are created by \LaTeX or BibTeX and which you don't need to delete or worry about:

example.bib – this is an important file that contains all the bibliographic information and references that you will be citing in the thesis for use with BibTeX. You can write it manually, but there are reference manager programs available that will create and manage it for you. Bibliographies in \LaTeX are a large subject and you may need to read about BibTeX before starting with this. Many modern reference managers will allow you to export your references in BibTeX format which greatly eases the amount of work you have to do.

MastersDoctoralThesis.cls – this is an important file. It is the class file that tells \LaTeX how to format the thesis.

main.pdf – this is your beautifully typeset thesis (in the PDF file format) created by \LaTeX . It is supplied in the PDF with the template and after you compile the template you should get an identical version.

main.tex – this is an important file. This is the file that you tell \LaTeX to compile to produce your thesis as a PDF file. It contains the framework and constructs that tell \LaTeX how to layout the thesis. It is heavily commented so you can read exactly what each line of code does and why it is there. After you put your own information into the *THESIS INFORMATION* block – you have now started your thesis!

Files that are *not* included, but are created by \LaTeX as auxiliary files include:

main.aux – this is an auxiliary file generated by \LaTeX , if it is deleted \LaTeX simply regenerates it when you run the main `.tex` file.

main.bbl – this is an auxiliary file generated by BibTeX, if it is deleted, BibTeX simply regenerates it when you run the main `.aux` file. Whereas the `.bib` file contains all the references

you have, this `.bbl` file contains the references you have actually cited in the thesis and is used to build the bibliography section of the thesis.

main.blg – this is an auxiliary file generated by BibTeX, if it is deleted BibTeX simply regenerates it when you run the main `.aux` file.

main.lof – this is an auxiliary file generated by L^AT_EX, if it is deleted L^AT_EX simply regenerates it when you run the main `.tex` file. It tells L^AT_EX how to build the *List of Figures* section.

main.log – this is an auxiliary file generated by L^AT_EX, if it is deleted L^AT_EX simply regenerates it when you run the main `.tex` file. It contains messages from L^AT_EX, if you receive errors and warnings from L^AT_EX, they will be in this `.log` file.

main.lot – this is an auxiliary file generated by L^AT_EX, if it is deleted L^AT_EX simply regenerates it when you run the main `.tex` file. It tells L^AT_EX how to build the *List of Tables* section.

main.out – this is an auxiliary file generated by L^AT_EX, if it is deleted L^AT_EX simply regenerates it when you run the main `.tex` file.

So from this long list, only the files with the `.bib`, `.cls` and `.tex` extensions are the most important ones. The other auxiliary files can be ignored or deleted as L^AT_EX and BibTeX will regenerate them.

7.5. Filling in Your Information in the `main.tex` File

You will need to personalise the thesis template and make it your own by filling in your own information. This is done by editing the `main.tex` file in a text editor or your favourite LaTeX environment.

Open the file and scroll down to the third large block titled *THESIS INFORMATION* where you can see the entries for *University Name*, *Department Name*, etc ...

Fill out the information about yourself, your group and institution. You can also insert web links, if you do, make sure you use the full URL, including the `http://` for this. If you don't want these to be linked, simply remove the `\href{url}{name}` and only leave the name.

When you have done this, save the file and recompile `main.tex`. All the information you filled in should now be in the PDF, complete with web links. You can now begin your thesis proper!

7.6. The `main.tex` File Explained

The `main.tex` file contains the structure of the thesis. There are plenty of written comments that explain what pages, sections and formatting the L^AT_EX code is creating. Each major document element is divided into commented blocks with titles in all capitals to make it obvious what the following bit of code is doing. Initially there seems to be a lot of L^AT_EX code, but this is all formatting, and it has all been taken care of so you don't have to do it.

Begin by checking that your information on the title page is correct. For the thesis declaration, your institution may insist on something different than the text given. If this is the case, just replace what you see with what is required in the *DECLARATION PAGE* block.

Then comes a page which contains a funny quote. You can put your own, or quote your favourite scientist, author, person, and so on. Make sure to put the name of the person who you took the quote from.

Following this is the abstract page which summarises your work in a condensed way and can almost be used as a standalone document to describe what you have done. The text you write will cause the heading to move up so don't worry about running out of space.

Next come the acknowledgements. On this page, write about all the people who you wish to thank (not forgetting parents, partners and your advisor/supervisor).

The contents pages, list of figures and tables are all taken care of for you and do not need to be manually created or edited. The next set of pages are more likely to be optional and can be deleted since they are for a more technical thesis: insert a list of abbreviations you have used in the thesis, then a list of the physical constants and numbers you refer to and finally, a list of mathematical symbols used in any formulae. Making the effort to fill these tables means the reader has a one-stop place to refer to instead of searching the internet and references to try and find out what you meant by certain abbreviations or symbols.

The list of symbols is split into the Roman and Greek alphabets. Whereas the abbreviations and symbols ought to be listed in alphabetical order (and this is *not* done automatically for you) the list of physical constants should be grouped into similar themes.

The next page contains a one line dedication. Who will you dedicate your thesis to?

Finally, there is the block where the chapters are included. Uncomment the lines (delete the % character) as you write the chapters. Each chapter should be written in its own file and put into the *Chapters* folder and named *LatexChapter*, *Chapter2*, etc. . . Similarly for the appendices, uncomment the lines as you need them. Each appendix should go into its own file and placed in the *Appendices* folder.

After the preamble, chapters and appendices finally comes the bibliography. The bibliography style (called *authoryear*) is used for the bibliography and is a fully featured style that will even include links to where the referenced paper can be found online. Do not underestimate how grateful your reader will be to find that a reference to a paper is just a click away. Of course, this relies on you putting the URL information into the BibTeX file in the first place.

7.7. Thesis Features and Conventions

To get the best out of this template, there are a few conventions that you may want to follow.

One of the most important (and most difficult) things to keep track of in such a long document as a thesis is consistency. Using certain conventions and ways of doing things (such as using a Todo list) makes the job easier. Of course, all of these are optional and you can adopt your own method.

7.7.1. Printing Format

This thesis template is designed for double sided printing (i.e. content on the front and back of pages) as most theses are printed and bound this way. Switching to one sided printing is as simple as uncommenting the *oneside* option of the `documentclass` command at the top of the `main.tex` file. You may then wish to adjust the margins to suit specifications from your institution.

The headers for the pages contain the page number on the outer side (so it is easy to flick through to the page you want) and the chapter name on the inner side.

The text is set to 11 point by default with single line spacing, again, you can tune the text size and spacing should you want or need to using the options at the very start of `main.tex`. The spacing can be changed similarly by replacing the *singlespacing* with *onehalfspacing* or *doublespacing*.

7.7.2. Using US Letter Paper

The paper size used in the template is A4, which is the standard size in Europe. If you are using this thesis template elsewhere and particularly in the United States, then you may have to change the A4 paper size to the US Letter size. This can be done in the margins settings section in `main.tex`.

Due to the differences in the paper size, the resulting margins may be different to what you like or require (as it is common for institutions to dictate certain margin sizes). If this is the case, then the margin sizes can be tweaked by modifying the values in the same block as where you set the paper size. Now your document should be set up for US Letter paper size with suitable margins.

7.7.3. References

The `bibtex` package is used to format the bibliography and inserts references such as this one [Reference1]. The options used in the `main.tex` file mean that the in-text citations of references are formatted with the author(s) listed with the date of the publication. Multiple references are separated by semicolons (e.g. [Reference2, Reference1]) and references with more than three authors only show the first author with *et al.* indicating there are more authors (e.g. [Reference3]). This is done automatically for you. To see how you use references, have a look at the `LatexChapter.tex` source file. Many reference managers allow you to simply drag the reference into the document as you type.

Scientific references should come *before* the punctuation mark if there is one (such as a comma or period). The same goes for footnotes¹. You can change this but the most important thing is to keep the convention consistent throughout the thesis. Footnotes themselves should be full, descriptive sentences (beginning with a capital letter and ending with a full stop). The APA6 states: «Footnote numbers should be superscripted, [...], following any punctuation mark except a dash.» The Chicago manual of style states: «A note number should be placed at the end of a sentence or clause. The number follows any punctuation mark except the dash, which it precedes. It follows a closing parenthesis.»

The bibliography is typeset with references listed in alphabetical order by the first author's last name. This is similar to the APA referencing style. To see how \LaTeX typesets the bibliography, have a look at the very end of this document (or just click on the reference number links in in-text citations).

A Note on `bibtex`

The `bibtex` backend used in the template by default does not correctly handle unicode character encoding (i.e. international characters). You may see a warning about this in the compilation log and, if your references contain unicode characters, they may not show up correctly or at all. The solution to this is to use the `biber` backend instead of the outdated

¹Such as this footnote, here down at the bottom of the page.

CUADRO 7.1: The effects of treatments X and Y on the four groups studied.

Groups	Treatment X	Treatment Y
1	0.2	0.8
2	0.17	0.7
3	0.24	0.75
4	0.68	0.3

bibtex backend. This is done by finding this in `main.tex`: `backend=bibtex` and changing it to `backend=biber`. You will then need to delete all auxiliary BibTeX files and navigate to the template directory in your terminal (command prompt). Once there, simply type `biber main` and `biber` will compile your bibliography. You can then compile `main.tex` as normal and your bibliography will be updated. An alternative is to set up your LaTeX editor to compile with `biber` instead of `bibtex`, see here for how to do this for various editors.

7.7.4. Tables

Tables are an important way of displaying your results, below is an example table which was generated with this code:

```
\begin{table}
\caption{The effects of treatments X and Y on the four groups studied.}
\label{tab:treatments}
\centering
\begin{tabular}{l l l}
\toprule
\tabhead{Groups} & \tabhead{Treatment X} & \tabhead{Treatment Y} \\
\midrule
1 & 0.2 & 0.8 \\
2 & 0.17 & 0.7 \\
3 & 0.24 & 0.75 \\
4 & 0.68 & 0.3 \\
\bottomrule
\end{tabular}
\end{table}
```

You can reference tables with `\ref{<label>}` where the label is defined within the table environment. See `LatexChapter.tex` for an example of the label and citation (e.g. Table 7.1).

7.7.5. Figures

There will hopefully be many figures in your thesis (that should be placed in the *Figures* folder). The way to insert figures into your thesis is to use a code template like this:

```
\begin{figure}
\centering
\includegraphics{Figures/Electron}
\decoRule
\caption[An Electron]{An electron (artist's impression).}
\label{fig:Electron}
\end{figure}
```

Also look in the source file. Putting this code into the source file produces the picture of the electron that you can see in the figure below.



FIGURA 7.1: An electron (artist's impression).

Sometimes figures don't always appear where you write them in the source. The placement depends on how much space there is on the page for the figure. Sometimes there is not enough room to fit a figure directly where it should go (in relation to the text) and so \LaTeX puts it at the top of the next page. Positioning figures is the job of \LaTeX and so you should only worry about making them look good!

Figures usually should have captions just in case you need to refer to them (such as in Figure 7.1). The `\caption` command contains two parts, the first part, inside the square brackets is the title that will appear in the *List of Figures*, and so should be short. The second part in the curly brackets should contain the longer and more descriptive caption text.

The `\decoRule` command is optional and simply puts an aesthetic horizontal line below the image. If you do this for one image, do it for all of them.

\LaTeX is capable of using images in pdf, jpg and png format.

7.7.6. Typesetting mathematics

If your thesis is going to contain heavy mathematical content, be sure that \LaTeX will make it look beautiful, even though it won't be able to solve the equations for you.

The «Not So Short Introduction to \LaTeX » (available on CTAN) should tell you everything you need to know for most cases of typesetting mathematics. If you need more information, a much more thorough mathematical guide is available from the AMS called, «A Short Math Guide to \LaTeX » and can be downloaded from: <ftp://ftp.ams.org/pub/tex/doc/amsmath/short-math-guide.pdf>

There are many different \LaTeX symbols to remember, luckily you can find the most common symbols in The Comprehensive \LaTeX -Symbol List.

You can write an equation, which is automatically given an equation number by \LaTeX like this:

```
\begin{equation}
E = mc^2
\label{eqn:Einstein}
\end{equation}
```

This will produce Einstein’s famous energy-matter equivalence equation:

$$E = mc^2 \tag{7.1}$$

All equations you write (which are not in the middle of paragraph text) are automatically given equation numbers by \LaTeX . If you don’t want a particular equation numbered, use the unnumbered form:

```
\[ a^2=4 \]
```

7.8. Sectioning and Subsectioning

You should break your thesis up into nice, bite-sized sections and subsections. \LaTeX automatically builds a table of Contents by looking at all the `\chapter{}`, `\section{}` and `\subsection{}` commands you write in the source.

The Table of Contents should only list the sections to three (3) levels. A `\chapter{}` is level zero (0). A `\section{}` is level one (1) and so a `\subsection{}` is level two (2). In your thesis it is likely that you will even use a `\subsubsection{}`, which is level three (3). The depth to which the Table of Contents is formatted is set within `MastersDoctoralThesis.cls`. If you need this changed, you can do it in `main.tex`.

7.9. In Closing

You have reached the end of this mini-guide. You can now rename or overwrite this pdf file and begin writing your own `LatexChapter.tex` and the rest of your thesis. The easy work of setting up the structure and framework has been taken care of for you. It’s now your job to fill it out!

Good luck and have lots of fun!

Guide written by —
 Sunil Patel: www.sunilpatel.co.uk
 Vel: LaTeXTemplates.com

Apéndice A

Frequently Asked Questions

A.1. How do I change the colors of links?

The color of links can be changed to your liking using:

```
\hypersetup{urlcolor=red}, or
```

```
\hypersetup{citecolor=green}, or
```

```
\hypersetup{allcolor=blue}.
```

If you want to completely hide the links, you can use:

```
\hypersetup{allcolors=.}, or even better:
```

```
\hypersetup{hidelinks}.
```

If you want to have obvious links in the PDF but not the printed text, use:

```
\hypersetup{colorlinks=false}.
```


Bibliografía

- [1] F Carminati y col. «Three dimensional Generative Adversarial Networks for fast simulation». En: *Journal of Physics: Conference Series* 1085.3 (2018), pág. 032016. ISSN: 1742-6596. DOI: 10.1088/1742-6596/1085/3/032016. URL: <https://iopscience.iop.org/article/10.1088/1742-6596/1085/3/032016><https://iopscience.iop.org/article/10.1088/1742-6596/1085/3/032016/meta>.
- [2] Federico Carminati y col. «Generative Adversarial Networks for fast simulation». En: *Journal of Physics: Conference Series* 1525.1 (2020), pág. 012064. ISSN: 1742-6596. DOI: 10.1088/1742-6596/1525/1/012064. URL: <https://iopscience.iop.org/article/10.1088/1742-6596/1525/1/012064><https://iopscience.iop.org/article/10.1088/1742-6596/1525/1/012064/meta>.
- [3] Peter Elmer, Mark Neubauer y Michael D. Sokoloff. «Strategic Plan for a Scientific Software Innovation Institute (S2I2) for High Energy Physics». En: (2017). arXiv: 1712.06592. URL: <https://arxiv.org/abs/1712.06592v2>.
- [4] Ralph Engel, Dieter Heck y Tanguy Pierog. «Extensive Air Showers and Hadronic Interactions at High Energy». En: <http://dx.doi.org/10.1146/annurev.nucl.012809.104544> 61 (2011), págs. 467-489. ISSN: 01638998. DOI: 10.1146/ANNUREV.NUCL.012809.104544. URL: <https://www.annualreviews.org/doi/abs/10.1146/annurev.nucl.012809.104544>.
- [5] M. Erdmann, J. Glombitza y D. Walz. «A deep learning-based reconstruction of cosmic ray-induced air showers». En: *Astroparticle Physics* 97 (2018), págs. 46-53. ISSN: 0927-6505. DOI: 10.1016/J.ASTROPARTPHYS.2017.10.006. URL: <https://www.sciencedirect.com/science/article/pii/S0927650517302219>.
- [6] Martin Erdmann, Jonas Glombitza y Thorben Quast. «Precise Simulation of Electromagnetic Calorimeter Showers Using a Wasserstein Generative Adversarial Network». En: *Computing and Software for Big Science* 2019 3:1 3.1 (2019), págs. 1-13. ISSN: 2510-2044. DOI: 10.1007/S41781-018-0019-7. URL: <https://link.springer.com/article/10.1007/s41781-018-0019-7>.
- [7] Martin Erdmann y col. «Generating and refining particle detector simulations using the Wasserstein distance in adversarial networks». En: *Computing and Software for Big Science* 2.1 (2018). arXiv: 1802.03325. URL: <https://arxiv.org/abs/1802.03325v1>.
- [8] David Foster. *Generative deep learning. Teaching machines to paint, write, compose, and play*. O'Reilly Media, Inc., 2019., 2019. ISBN: 9781492041948.
- [9] Jonas Glombitza. «Air-Shower Reconstruction at the Pierre Auger Observatory based on Deep Learning». En: *PoS ICRC2019* (2020), pág. 270. DOI: 10.22323/1.358.0270. URL: http://www.auger.org/archive/authors_icrc_2019.html<https://inspirehep.net/literature/1819455>.
- [10] Peter K.F. Grieder. *Extensive Air Showers. High Energy Phenomena and Astrophysical Aspects - A Tutorial Reference Manual and Data Book*. Springer Berlin, Heidelberg, 2010. ISBN: 978-3-540-76941-5. DOI: <https://doi.org/10.1007/978-3-540-76941-5>.
- [11] Michela Paganini, Luke de Oliveira y Benjamin Nachman. «Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multi-Layer Calorimeters». En: *Physical Review Letters* 120.4 (2017). DOI: 10.1103/PhysRevLett.120.042003. arXiv: 1705.02355v2. URL: <http://arxiv.org/abs/1705.02355><http://dx.doi.org/10.1103/PhysRevLett.120.042003>.

-
- [12] Michela Paganini, Luke de Oliveira y Benjamin Nachman. «CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks». En: *Physical Review D* 97.1 (2017). DOI: 10 . 1103 / physrevd . 97 . 014021. arXiv: 1712 . 10321. URL: <https://arxiv.org/abs/1712.10321v1>.
- [13] Piergiorgio Picozza y V. Malvezzi. *Cosmic Rays under the Knee*. 2008. URL: https://www.researchgate.net/publication/265671393_Cosmic_Rays_under_the_Knee (visitado 12-05-2022).