



UNIVERSIDAD
NACIONAL
DE LA PLATA

FACULTAD DE INFORMÁTICA

TESINA DE LICENCIATURA

TÍTULO: Estudio de métodos y técnicas de aprendizaje por transferencia en el contexto de aprendizaje automático

AUTORES: Suárez, José Manuel

DIRECTOR: Dra. Pons, Claudia

CODIRECTOR: Lic. Pérez, Gabriela

ASESOR PROFESIONAL:

CARRERA: Licenciatura en Informática

Resumen

La presente tesina de grado de investigación teórica propone un recorrido por los fundamentos que giran en torno a la técnica de aprendizaje por transferencia, sub-áreas y metodologías resolutivas presentes en distintos escenarios aplicativos, así como proyecciones de trabajos futuros para el desarrollo en el área.

Palabras Clave

Aprendizaje a largo plazo, Aumentación de datos, Aplicaciones de TL, Enfoques de transferencia, Extracción de características, Reducción de dimensionalidad, Selección de características, Transferencia de conocimiento, Técnicas de transferencia, Transferencia negativa.

Conclusiones

Se expuso el potencial esgrimido por el TL como justificativo de su acelerado desarrollo sobre disciplinas multivariadas y recorriendo técnicas complementarias que explotan aún más su potencial. Permitiéndole al lector, comprender tópicos fundamentales a fin de discernir e implantar las metodologías más adecuadas ante cada problemática afrontada.

Trabajos Realizados

Para el despliegue de esta tesina se realizaron tareas de búsqueda, investigación, análisis, recopilación y reflexión, sobre variadas fuentes informativas como papers, artículos, noticias y publicaciones relacionadas. Analizando problemáticas del ML que impulsan distintas sub-áreas del TL, ilustrando su aplicación sobre un caso de estudio realista propuesto.

Trabajos Futuros

Este trabajo describe problemáticas de relevancia y resolución actualmente abierta, por tanto, el abordaje de cualquiera de estos tópicos resultaría de gran valor para el ML. En relación al caso de estudio planteado y el carácter teórico investigativo de la tesina, concretar la implementación propuesta así como el análisis de los resultados obtenidos a fin de comprobar su congruencia con las conclusiones expresadas aquí, o incluso mejor, la promoción de novedosos métodos superadores.

Fecha de la presentación: Junio de 2021

Tesina de grado

Estudio de métodos y técnicas de aprendizaje por transferencia en el contexto de aprendizaje automático

Autor
Suárez, José Manuel

Director
Dra. Pons, Claudia

Codirector
Lic. Pérez, Gabriela

Titulación
Licenciatura en Informática

Lugar y fecha de presentación
La Plata - Junio de 2021



Facultad de Informática
Universidad Nacional de La Plata (UNLP)

Resumen

La presente tesina de grado de investigación teórica propone un recorrido por los fundamentos que giran en torno a la técnica de aprendizaje por transferencia, subáreas y metodologías resolutivas presentes en distintos escenarios aplicativos, así como proyecciones de trabajos futuros para el desarrollo en el área.

Palabras clave

Aprendizaje a largo plazo, Aumentación de datos, Aplicaciones de TL, Enfoques de transferencia, Extracción de características, Reducción de dimensionalidad, Selección de características, Transferencia de conocimiento, Técnicas de transferencia, Transferencia negativa.

Trabajos realizados

Para el despliegue de esta tesina se realizaron tareas de búsqueda, investigación, análisis, recopilación y reflexión, sobre variadas fuentes informativas como papers, artículos, noticias y publicaciones relacionadas. Analizando problemáticas del **ML** que impulsan distintas subáreas del **TL**, ilustrando su aplicación sobre un caso de estudio realista propuesto.

Conclusiones

Se expuso el potencial esgrimido por el **TL** como justificativo de su acelerado desarrollo sobre disciplinas multivariadas y recorriendo técnicas complementarias que explotan aún más su potencial. Permitiéndole al lector, comprender tópicos fundamentales a fin de discernir e implantar las metodologías más adecuadas ante cada problemática afrontada.

Trabajos futuros

Este trabajo describe problemáticas de relevancia y resolución actualmente abierta, por tanto, el abordaje de cualquiera de estos tópicos resultaría de gran valor para el **ML**. En relación al caso de estudio planteado y el carácter teórico investigativo de la tesina, concretar la implementación propuesta así como el análisis de los resultados obtenidos a fin de comprobar su congruencia con las conclusiones expresadas aquí, o incluso mejor, la promoción de novedosos métodos superadores.

Agradecimientos

En el plano educativo, agradezco a todos los que han luchado por el acceso a la educación pública, gratuita y de calidad, gracias a la cuál me formé. A la **UNLP**, que me ha constituido como profesional y persona; más precisamente a la facultad de informática y **LIFIA** por su apoyo fundamental que posibilitó la finalización de esta tesina. A la comunidad de profesores, por su vocación y dedicación hacia la comunidad estudiantil, y en un sentido especial a Claudia Pons, por su interés y acompañamiento constante.

En el plano personal, a Luján Marconi, por su amor ilimitado y ser mi firme compañera en cualquier circunstancia. A mi familia, por inculcarme el valor del conocimiento, el esfuerzo por alcanzarlo y perfeccionarlo constantemente, reforzado por la integridad y humildad. A Verónica Tarrío y Diego Tarrío, por instarme a estudiar informática y su apoyo fundamental en etapas iniciales de mi carrera. A Laura Suárez, por su ayuda y profesionalismo que elevaron el nivel de este trabajo.

Ante todo agradezco a Dios, quien me ha regalado todo por Amor.

Índice general de contenidos

Índice de figuras	8
Índice de tablas	8
Índice de abreviaturas y siglas	9
Índice de palabras clave	12
Capítulo 1 - Introducción	14
1.1 Contexto	14
1.2 Motivación	15
1.3 Objetivos	16
1.3.1 Objetivos generales	16
1.3.2 Objetivos específicos	16
1.4 Metodología de trabajo	16
1.5 Resultados esperados	16
1.6 Estructura de la tesina	17
1.7 Conclusión	18
Capítulo 2 - Aprendizaje por transferencia	20
2.1 Introducción	20
2.2 Concepto	20
2.3 Motivación	21
2.4 Objetivo	22
2.5 Definiciones	24
2.6 Técnicas	25
2.6.1 Transferencia inductiva	25
2.6.2 Transferencia transductiva	26
2.6.3 Transferencia no supervisada	26
2.6.4 Transferencia por refuerzo	27
2.7 Enfoques	28
2.7.1 Transferencia de instancias	29
2.7.2 Transferencia de representación de características	30
2.7.3 Transferencia de parámetros	30
2.7.4 Transferencia de conocimiento relacional	31
2.8 Limitaciones	31
2.9 Aprendizaje automático de largo plazo	32
2.10 Conclusión	33

Capítulo 3 - Estado del arte	35
3.1	Introducción 35
3.2	Escenarios motivacionales para TL 35
3.2.1	Clasificación de páginas web sin ejemplos negativos 35
3.2.2	TL para localización en interiores basada en señales WiFi 36
3.2.3	Adaptación de dominios para la clasificación de sentimientos 37
3.3	TrAdaBoost: un impulso para TL 39
3.4	Aprendizaje autodidacta: TL desde datos sin etiquetar 41
3.5	TL mediante mapeo estructural local de modelos múltiples 43
3.6	Mapeo y revisión de redes lógicas de Markov para TL 45
3.7	Aprendizaje y evaluación de clasificadores bajo el sesgo de selección de la muestra 48
3.8	Aprendizaje por transferencia mediante reducción de dimensionalidad. 50
3.9	Reducción de dimensionalidad transferida 52
3.10	Transferencia entre dominios para el aprendizaje por refuerzo 55
3.11	Distant Domain Transfer Learning 57
3.12	Conclusión 58
Capítulo 4 - Propuesta metodológica para la aplicación de TL	60
4.1	Introducción 60
4.2	Etapas del proceso resolutivo 60
4.3	Flujo metodológico 63
4.3.1	Profundización de etapa de consideración e implantación de técnicas de TL 69
4.4	Conclusión 72
Capítulo 5 - Un caso aplicativo	74
5.1	Introducción 74
5.2	Presentación del modelo y las premisas 74
5.3	Problemática afrontada. 75
5.4	Solución propuesta 76
5.5	Conclusión 81
Capítulo 6 - Aportes, conclusiones y trabajos futuros	83
6.1	Introducción 83
6.2	Aportes 83
6.3	Conclusiones 84
6.4	Trabajos futuros 84
6.4.1	Líneas de investigación propuestas 87
6.5	Conclusión 88

Anexos		90
A	Aumentación de datos	90
	A.1 Origen y definición del problema	90
	A.2 Técnicas	91
B	Reducción de dimensionalidad	92
	B.1 Origen y definición del problema	92
	B.2 La maldición de la dimensionalidad	93
	B.3 Técnicas y metodologías	95
<u>C</u>	Aplicaciones del aprendizaje por transferencia	98
	C.1 Principales aplicaciones y desarrollos	98
	C.1.1 Agrupamiento y clasificación de texto	98
	C.1.2 Agrupamiento y clasificación de imágenes	99
	C.1.3 Aprendizaje reforzado	99
	C.1.4 Clasificación sentimental	100
	C.1.5 Filtrado colaborativo	100
	C.1.6 Estimación de ubicación basada en sensores	101
	C.1.7 Aprendizaje de métricas	102
Glosario		104
Referencias bibliográficas		115

Índice de figuras

Figura 2.1: Diferencias entre el proceso de aprendizaje en ML tradicional y TL.	20
Figura 2.2: Fuentes de información en el TL.	21
Figura 2.3: Factores de mejora de la transferencia respecto del aprendizaje	23
Figura 2.4: Diagrama integral de técnicas y metodologías de TL	27
Figura 2.5: Interacción de un agente de aprendizaje por refuerzo con su entorno	28
Figura 2.6: Métodos de TL y su relación con el rendimiento y grado de afinidad entre tareas.	32
Figura 3.1: Precisión de métodos para la adaptación de dominios	38
Figura 3.2: Esquema funcional del marco de trabajo TrAdaBoost	40
Figura 3.3: Distinción entre paradigmas según datos etiquetados	42
Figura 3.4: Bases aprendidas por SCA para una imagen	43
Figura 3.5: Tasas de error con datos de entrenamiento sesgados e insesgados	49
Figura 3.6: Comparación de precisión en métodos asistidos por MMDE	51
Figura 3.7: Desempeño de métodos PCA+K-medias, DisK-medias y TDA	53
Figura 3.8: Curvas de transferencia de aprendizaje hacia MAle	56
Figura 3.9: Proceso de transferencia en SLA con dominios intermedios	57
Figura 4.1: Etapas resolutorias en un proyecto de ML.	63
Figura 4.2: Diagrama de flujo de procesos en un proyecto de TL	68
Figura 4.3: Diagrama de flujo de procesos del “Análisis de técnicas y enfoques” de TL	71
Figura A.1: Tasas de error de entrenamiento y prueba (sobreajuste vs. generalización)	90
Figura B.1: Proceso de tratamiento de datos de alta dimensionalidad.	93
Figura B.2: Relación entre dimensionalidad y dispersión de los datos.	93
Figura B.3: Relación entre dimensionalidad y rendimiento en modelos de ML.	94
Figura B.4: Esquema de métodos de reducción de dimensionalidad	97

Índice de tablas

Tabla 2.1: Relación entre ML tradicional y técnicas de TL	25
Tabla 2.2: Técnicas de TL y disponibilidad de datos etiquetados en dominios	27
Tabla 2.3: Técnicas de TL y enfoques utilizados	29
Tabla 3.1: Comparación de rendimiento de métodos en diferentes conjuntos de datos	45
Tabla 3.2: Métricas TR y PI en términos de AUC y CLL sobre ScrKD	47
Tabla 3.3: Tiempos de entrenamiento y construcción de mapeos	47
Tabla 3.4: Métodos de clasificación y su afectación ante el SSM	48
Tabla 3.5: Precisión de clasificadores entrenados sobre diferentes espacios de características.	51
Tabla 3.6: Resultados sobre el conjunto de datos de AT&T con métrica ACC	54

Índice de abreviaturas y siglas

A

AAAI	Asociación para el avance de la inteligencia artificial (A ssociation for the A dvancement of A rtificial I ntelligence)
ANN	Red neuronal artificial (A rtificial N eural N etwork)
AP	Punto de acceso (A ccess P oint)
AUC	Área bajo la curva de recuperación de precisión (A rea U nder the precision-recall C urve)

C

CF	Filtrado colaborativo (C ollaborative F iltering)
CLL	Probabilidad de registro condicional (C onditional L og- L ikelihood)
CNN	Red neuronal convolucional (C onvolutional N eural N etwork)
CoD	Maldición de dimensionalidad (C urse of D imensionality)
CS	Cambio covariante (C ovariate S hift)
CST	Sistema de transferencia coordinado (C oordinate S ystem T ransfer)

D

Data+	Aumentación de datos (D ata A ugmentation)
DA	Adaptación de dominios (D omain A daptation)
DDTL	Transferencia entre dominios distantes (D istant D omain T ransfer L earning)
DIA	Área de diagnóstico por imágenes (D iagnostic I maging A rea)
DL	Aprendizaje profundo (D eep L earning)
DLM	Modelo de aprendizaje profundo (D eep L earning M odel)
DR	Reducción de dimensionalidad (D imensionality R eduction)
DTA	Algoritmo de árbol de decisión (D ecision T ree A lgorithm)

E

ECML	Conferencia europea sobre aprendizaje automático (E uropean C onference on M achine L earning)
EDA	Análisis exploratorio de datos (E xploratory D ata A nalysis)
ERM	Minimización del riesgo empírico (E mpirical R isk M inimization)

F

FE	Extracción de características (F eature E xtraction)
FEA	Algoritmos de extracción de características (F eature E xtraction A lgorithms)
FS	Selección de características (F eature S election)
FSA	Algoritmos de selección de características (F eature S election A lgorithms)
FORTE	Revisión de teorías de primer orden desde ejemplos (F irst- O rders R evisión of T heories from E xamples)

G

GAN	Red generativa adversaria (Generative Adversarial Network)
GDA	Análisis de discriminante Gaussiano (Gaussian Discriminant Analysis)
GP	Proceso Gaussiano (Gaussian Process)

H

HCA	Área del cuidado de la salud (Health Care Area)
HetTL	Aprendizaje por transferencia heterogéneo (Heterogeneous Transfer Learning)
HomTL	Aprendizaje por transferencia homogéneo (Homogeneous Transfer Learning)

I

IA	Inteligencia artificial (Artificial Intelligence)
ICDM	Conferencia internacional sobre minería de datos (International Conference on Data Mining)
ICML	Conferencia internacional sobre aprendizaje automático (International Conference on Machine Learning)
IEEE	Instituto de ingenieros eléctricos y electrónicos (Institute of Electrical and Electronics Engineers)
IID	Independiente e idénticamente distribuída (Independent and Identically Distributed)
ILP	Programación lógica inductiva (Inductive Logic Programming)
IoT	Internet de las cosas (Internet of Things)
IT	Tecnologías de la información (Information Technology)
ITL	Aprendizaje por transferencia inductivo (Inductive Transfer Learning)
IVM	Máquina de vector informativo (Informative Vector Machine)

K

KB	Base de conocimiento (Knowledge Base)
KNNa	Algoritmo de K vecinos más cercanos (K-nearest Neighbours algorithm)

L

LapRLSR	Regresor Laplaciano de mínimos cuadrados regularizado (Laplacian Regularized Least Square Regressor)
LDA	Análisis de discriminante lineal (Linear Discriminant Analysis)
LIFIA	Laboratorio de Investigación y Formación en Informática Avanzada
LML	Aprendizaje automático para toda la vida (Lifelong Machine Learning)
LPP	Proyecciones preservadoras de localidad (Locality Preserving Projections)
LR	Regresión logística (Logistic Regression)
LWE	Marco de trabajo conjunto ponderado localmente (Locally Weighted Ensemble Framework)

M

MCa	Algoritmo de mapeo convergencia (Mapping Convergence algorithm)
MCF	Filtrado colaborativo multidominio (Multi-domain Collaborative Filtering)
MDP	Proceso de decisión de Markov (Markov Decision Process)
MIM	Método de información mutua (Mutual Information Method)

ML	Aprendizaje automático (M achine L earning)
MLN	Red lógica de Markov (M arkov L ogic N etwork)
MMDE	Método de Incrustación máxima de discrepancia media (M aximum M ean D iscrepancy E mbedding)
MSE	Error cuadrático medio (M ean S quared E rror)
MtL	Aprendizaje multitarea (M ulti-task L earning)
MVU	Despliegue de varianza máxima (M aximum V ariance U nfolding)
N	
NIPS	Sistema de procesamiento de información neuronal (N eural I nformation P rocessing S ystems)
NLP	Procesamiento de lenguaje natural (N atural L anguage P rocessing)
NT	Transferencia negativa (N egative T ransfer)
O	
OSVM	Máquina de vector de soporte de clase única (O ne-class S upport V ector M achine)
P	
PCAa	Algoritmo de análisis de componentes principales (P rincipal C omponent A nalysis algorithm)
PDDL	Lenguaje de descripción de dominio de planificación (P lanning D omain D escription L anguage)
PEBL	Aprendizaje basado en ejemplos positivos (P ositive E xample B ased L earning)
PKDD	Principios y prácticas del descubrimiento del conocimiento en bases de datos (P inciples and P ractice of K nowledge D iscovery in D atabases)
pLWE	Marco parcial conjunto ponderado localmente (p artial L ocally W eighted E nsemble)
PMF	Marco de factorización matricial probabilística (P robabilistic M atrix F actorization)
PT	Transferencia positiva (P ositive T ransfer)
R	
RFa	Algoritmo de bosque aleatorio (R andom F orest algorithm)
RL	Aprendizaje por refuerzo o reforzado (R einforcement L earning)
RLSR	Regresor de mínimos cuadrados regularizado (R egularized L east S quare R egressor)
RPF	Búsqueda del camino relacional (R elational P athfinding)
RSS	Intensidad de señal recibida (R eceived S ignal S trength)
RT	Método de transferencia de reglas (R ule T ransfer method)
RTL	Aprendizaje por transferencia reforzado (R einforcement T ransfer L earning)
S	
SCA	Algoritmo de codificación dispersa (S pase C oding A lgorithms)
SCL	Aprendizaje de correspondencia estructural (S tructural C orrespondence L earning)
SCL-MI	Método de correspondencia estructural de información mutua (S tructural C orrespondence L earning M utual I nformation)
SDL	Aprendizaje profundo supervisado (S upervised D eep L earning)
SLA	Algoritmo de aprendizaje selectivo (S elective L earning A lgorithm)

SMA	Marco de trabajo de promediado simple de modelo (S imple M odel A veraging framework)
SMSVM	Máquina de vectores de soporte de margen flexible (S oft M argin S upport V ector M achine)
SSB	Sesgo de selección de la muestra (S ample S election B ias)
SsL	Aprendizaje semi-supervisado (S emi-supervised L earning)
StC	Algoritmo de agrupamiento autodidacta (S elf-taught C lustering)
StL	Aprendizaje autodidacta (S elf-taught L earning)
STL	Aprendizaje por transferencia supervisado (S upervised T ransfer L earning)
SVM	Máquina de vectores de soporte (S upport V ector M achine)
SVR	Regresión de vectores de soporte (S upport V ector R egression)

T

TAMaR	Transferencia mediante mapeo y revisión automática (T ransfer via A utomatic M apping and R evision)
TDA	Algoritmo de análisis de discriminativo transferido (T ransferred D iscriminative A nalysis)
TDR	Reducción de dimensionalidad transferida (T ransferred D imensionality R eduction)
TL	Aprendizaje por transferencia (T ransfer L earning)
TSVM	Máquinas de vectores de soporte transductivas (T ransductive S upport V ector M achines)
TTL	Aprendizaje por transferencia transductivo (T ransductive T ransfer L earning)

U

UNLP	Universidad nacional de La Plata
UTL	Aprendizaje por transferencia no supervisado (U n supervised T ransfer L earning)

W

WiFi	Red electrónica de interconexión inalámbrica (W ireless F idelity)
WILP	Problema de localización en interiores basado en señales WiFi (W iFi-based I ndoor L ocalization P roblem)
WNNa	Algoritmo de arrojo (W innow algorithm)
WPLL	Probabilidad de registro pseudo ponderado (W eighted P seudo L og- L ikelihood)

Índice de palabras clave

Aplicaciones de TL	Reducción de dimensionalidad
Aprendizaje a largo plazo	Selección de características
Aumentación de datos	Técnicas de transferencia
Enfoques de transferencia	Transferencia de conocimiento
Extracción de características	Transferencia negativa.

Capítulo 1 - Introducción

En este trabajo se abordarán diferentes aspectos relacionados a la temática del aprendizaje por transferencia (**TL**), subárea de investigación correspondiente al aprendizaje automático (**ML**) y a su vez, perteneciente al ámbito de la inteligencia artificial (**AI**). Detallando las problemáticas que este novedoso enfoque apunta a resolver, así como subclasificaciones y metodologías más relevantes, objetivos, ventajas y desventajas de implantación.

En un aspecto personal, la investigación y desarrollo de la temática de transferencia de conocimiento en el contexto del ML, me resultó una tarea enriquecedora debido al carácter novedoso que representaba en lo relativo a mi perfil técnico, profesional y su relevancia actual. En los últimos años se ha visto potenciado el avance y masiva utilización de las redes neuronales dentro del contexto de la IA, entendiendo por tales motivos, que el tópico central de este trabajo (TL) representa el siguiente paso evolutivo a concretar en el avance del área.

El presente trabajo se enmarca dentro del área de la investigación teórica y si bien abordará un amplio abanico de enfoques y metodologías innovadoras, el mismo se encuentra focalizado no sólo a presentar al lector estos temas sino además a plasmarlo sobre un marco aplicativo general ejemplificado sobre un caso aplicativo concreto que permita ponderar la importancia que el TL esgrime dentro del ML. Entre las limitaciones de este trabajo podemos mencionar que si bien posee un nivel de profundidad significativo, acorde a un trabajo de esta categoría, no pretende incurrir en los pormenores de las metodologías y algoritmos abordados, lo cual requeriría el abordaje de problemas de alta complejidad matemática y estadística, que no representa el objetivo central de esta tesina ni el interés general del lector del área de **IT**. Sin embargo, y a fin de una profundización exhaustiva de esos temas, el presente trabajo facilita a los interesados abundante material citado en el **índice de referencias bibliográficas**.

1.1 - Contexto

El TL surge debido a la necesidad de solventar un conjunto de problemáticas que frecuentemente se suscitan en el contexto del ML o incluso para lograr un refinamiento de las soluciones existentes. El despliegue exitoso de técnicas de TL posibilita el perfeccionamiento del proceso de aprendizaje y, por ende, el aumento del desempeño de los modelos de ML construidos, con el objetivo de mitigar o directamente evitar el acontecimiento de situaciones como las mencionadas a continuación:

- Carencia de conjuntos de datos de dominio público y tamaño considerable, que se ajusten a una tarea objetivo particular.
- Conjuntos de datos de elevada dimensionalidad o escaso número de instancias.
- Conjuntos de datos de entrenamiento y pruebas pertenecientes a diferentes espacios de características o bajo distintas distribuciones de probabilidad subyacentes.
- Requerimientos de hardware muy demandantes y en ocasiones prohibitivos para el aprendizaje de una determinada tarea desde cero.
- Elevado costo y esfuerzo de recopilación, etiquetado, y preprocesamiento de los datos.

- Extensos períodos de tiempo y costos necesarios para el entrenamiento de modelos desde cero.
- Alto nivel de conocimientos demandados para la configuración de sistemas de ML desde cero.

La aplicación de la técnica de TL en el contexto del ML no sólo presenta ventajas significativas cuando es realizada adecuadamente (transferencia positiva, **PT**), sino que al mismo tiempo, puede acarrear severas desventajas al ser aplicada en casos indebidos o de forma incorrecta, pudiendo resultar en un rendimiento de los modelos resultantes incluso peor que aquel que se hubiese obtenido sin aplicar técnicas de TL (fenómeno conocido como transferencia negativa, **NT**). Por consiguiente, cobra un rol fundamental el estudio de las situaciones en las que es posible, recomendable o inapropiado aplicar TL. [1]

1.2 - Motivación

La elaboración de la presente tesina se encuentra motivada principalmente por el “florecimiento” que recientemente ha tenido el área del ML, especialmente potenciado por el desarrollo de las redes neuronales (**ANNs**) y su masiva utilización gracias a una combinación de elementos como el incremento de la capacidad de cómputo, algoritmos más sofisticados, el arribo del big data y la minería de datos, el acelerado incremento de dispositivos conectados en red (**IoT**) y requerimientos de acciones automatizadas. Este cambio radical potenciado en los últimos años, promete (y ya evidencia) un impacto sin precedentes en la ciencia, llegando a ser catalogado como una nueva revolución tecnológica. El TL, como rama del ML, ha sido traccionado por este aumento de investigaciones y desarrollos, así como el surgimiento de novedosos y variados ámbitos de aplicación, que lo convierte en un terreno muy atractivo para ser recorrido.

Se pretende desplegar un análisis completo y de considerable profundidad, describiendo motivaciones, conceptos fundamentales, recorrido de trabajos novedosos más sobresalientes, aspectos favorables y desfavorables presentes en cada metodología, recorriendo áreas de aplicación y expectativas a futuro. En síntesis, se tiene como objetivo fundamental la exploración y ampliación de conocimiento sobre un área prometedora y con abundante proyección futura, como así también se intentará complementar estos objetivos con la proposición de metodologías aplicativas y el análisis de casos concretos de estudio, a fin de incorporar un enfoque pragmático sobre este trabajo, que eleve su valor y utilidad aplicativa.

Asimismo, en un ámbito más personal, representa un tema completamente novedoso sobre el cuál no he tenido la oportunidad de incursionar previamente en el aspecto académico ni en el laboral, representando un área de sumo interés que como se ha mencionado, experimenta un sostenido desarrollo y promete en el corto y mediano plazo, acaparar buena parte de la industria de desarrollo.

1.3 - Objetivos

1.3.1 - Objetivos generales

Los objetivos generales que persigue esta investigación giran en torno a dos aspectos fundamentales, correspondientes al plano académico y contributivo:

- Incursionar en áreas de IA, ML y TL, profundizar en la investigación de estas temáticas, que han tomado recientemente carácter de ineludibles para cualquier profesional de ciencias informáticas.
- Contribuir a la comunidad académica y de IT, nutrir mi aspecto cooperativo y retributivo hacia la comunidad académica, anhelando que mi tesina sea tomada como base referencial por parte de estudiantes y profesionales de IT, impulsando el desarrollo de investigaciones novedosas.

1.3.2 - Objetivos específicos

Entre los objetivos específicos que persigue el desarrollo de esta investigación podemos mencionar: brindar una visión general de las problemáticas, motivaciones y objetivos del TL; proporcionar conceptos y definiciones fundamentales, necesarios para la comprensión de las metodologías abordadas; analizar las diferentes sub-áreas propias del TL, sus metodologías resolutivas, así como también la revisión de los trabajos más relevantes conjuntamente con el abordaje de sus respectivos contextos de aplicación; analizar las limitaciones o eventuales desventajas aplicativas que conllevan las técnicas de TL, además de enunciar las perspectivas y escenarios de aplicación futuros.

Brindar al lector de este trabajo información clara y precisa que le permita integrar los contenidos expuestos a través del desarrollo del marco teórico, con un marco metodológico general y su aplicación sobre un caso de estudio realista y actual, con el objetivo de disponer de una mirada más abarcativa y acabada de la técnica de TL, que posibilite discernir ante qué situaciones del mundo real puede aplicarse correctamente.

También se destinan secciones anexas a fin de profundizar en tópicos que, si bien no son requeridos para la comprensión de la línea de investigación principal, resultarán de utilidad en cuanto a la contextualización y repaso complementario, especialmente en lo referente al lector novato, es decir, aquel que recientemente comienza a incursionar en el área, cuya experiencia y lectura previa es limitada.

1.4 - Metodología de trabajo

La forma de trabajo implementada consiste básicamente de la búsqueda, recopilación, análisis y tratamiento de información certificada y validada, proveniente de organizaciones, personalidades o referentes, publicaciones y artículos académicos del ámbito del ML. Cada sección presentada contendrá su referencia bibliográfica debidamente indicada, además de enlaces externos cuando sean requeridos. Por su parte, las siglas utilizadas figuran apropiadamente descriptas en una **sección** propia y complementada con un **glosario** que a su vez puede ser profundizado mediante referencias externas en caso de ser necesario.

1.5 - Resultados esperados

Conocer y profundizar en los motivos de surgimiento y evolución de esta rama del ML, estableciendo un marco teórico de investigación a fin de afianzar pilares de conocimiento requeridos para quienes lean este trabajo.

Desarrollar el estado del arte del TL, mediante el análisis de una nutrida batería de metodologías más sobresalientes y novedosas a fin de ponderar la diversidad de escenarios aplicativos y mensurar su potencial.

Descubrir y comprender contextos de aplicación, problemáticas, ventajas y desventajas, presentes dentro de las diferentes subclasificaciones del TL. Confiriendo una visión amplia e integral del tema.

Establecer sólidos lineamientos que permitirán diferenciar las distintas áreas del TL y conocer bajo qué contextos se aplica cada una de ellas, en base a las características del conjunto de datos a utilizar.

Conocer en profundidad los aspectos negativos que pueden darse al aplicar el TL y aprender a identificar las características que determinan los contextos donde no es recomendable su implantación.

Describir un marco metodológico aplicativo de alto nivel de abstracción que pueda ser considerado previamente y resulte de utilidad para cualquiera que se aboque a la resolución de un determinado problema utilizando TL.

Contar con un caso aplicativo para hacer tangibles los conocimientos teóricos vistos y tener una noción más palpable del potencial del TL para resolver problemáticas que se presentan en el ML tradicional.

Presentar las más relevantes técnicas complementarias del TL en lo referente al pre-procesamiento y tratamiento de datos, que son requeridas para potenciar su rendimiento.

Por último, apreciar los múltiples y variados ámbitos de aplicación que pueden beneficiarse del TL, y al mismo tiempo delinear los desafíos por afrontar y superar a futuro en lo relativo al ML.

1.6 - Estructura de la tesina

El presente trabajo se encuentra organizado en capítulos que mantienen una estructura homogénea, describiendo secciones de introducción, desarrollo y conclusión de cada temática puntual abordada.

Capítulo 1 - Introducción

Se presenta el prólogo de la tesina, donde se aborda el contexto del trabajo, motivaciones y objetivos perseguidos, metodología de investigación y desarrollo utilizada, resultados deseados y estructura organizativa de la misma.

Capítulo 2 - Aprendizaje por transferencia

Se presenta el marco teórico de la tesina, brindando conceptos y definiciones, contextos de implantación y clasificaciones asociadas al TL, revisando implicaciones, objetivos, ventajas y desventajas aplicativos. Se aborda el problema de la transferencia negativa, factores propiciatorios y herramientas de tratamiento. Además se incursiona en un tipo de transferencia de largo plazo muy relevante.

Capítulo 3 - Estado del arte

Se presentan y analizan diferentes investigaciones y trabajos asociados a la temática del TL, que representan algunos de los avances más novedosos en distintas áreas fundamentales, erigiéndose como hitos investigativos que orientan las trayectorias de avance de cada subárea del TL.

Capítulo 4 - Propuesta metodológica para la aplicación de TL

Se presenta una descripción detallada de las dificultades que pueden presentarse en cada paso propio del proceso de desarrollo de ML, exponiendo las mejores y más apropiadas alternativas resolutorias dentro del área del TL y finalmente se consolida ese análisis en un flujo de procesos.

Capítulo 5 - Un caso aplicativo

Se presenta un enfoque práctico de los contenidos vistos, a partir del análisis de un caso de estudio concreto que inicia con el planteamiento de la problemática abordada y determinadas premisas de difícil cumplimiento, intentando evidenciar de qué forma la aplicación de técnicas de TL permiten lograr los objetivos propuestos y superar las dificultades que se presentan bajo un escenario realista.

Capítulo 6 - Aportes, conclusiones y trabajos futuros

Se enumeran los aportes realizados por esta tesina, las conclusiones a las que se ha arribado a partir de su desarrollo, así como la mención de las distintas áreas de aplicabilidad futura y trabajos de investigación que pueden desencadenarse a partir de este trabajo.

Anexo A - Aumentación de datos

Se presenta un tema fundamental para afrontar el problema que representa la carencia de datos etiquetados en la construcción de modelos de ML, abordando las metodologías principales ya que conocerlas permitirá establecer las herramientas disponibles para solventar problemas de este tipo.

Anexo B - Reducción de dimensionalidad

Se presenta un tema de suma importancia en aplicaciones donde la dimensionalidad de los datos es elevada, abordando en profundidad el surgimiento del problema así como las metodologías existentes para su resolución. Su abordaje posibilitará al lector disponer de conocimientos sólidos para lograr modelos de ML más ligeros y que trabajen sobre conjuntos minimalistas de datos relevantes.

Anexo C - Aplicaciones del aprendizaje por transferencia

Se estudian diferentes aplicaciones asociadas al TL, así como su impacto sobre cada área y los aspectos más innovadores de cada una. Conocerlas será de suma utilidad para hacerse una representación concreta del potencial y amplio espectro aplicativo que las técnicas de TL poseen.

1.7 - Conclusión

En este capítulo, se ha propuesto una introducción al TL dentro del campo del ML, analizando contexto, problemáticas y motivaciones que potenciaron su surgimiento, nutriendo de manera constante su desarrollo y evolución. Además se han presentado conceptos fundamentales, estableciendo objetivos generales y específicos que se persiguen con la realización de esta tesina, descripción de la metodología de trabajo implementada, resultados esperados, así como organización estructural que la misma presenta.

El objetivo de este capítulo introductorio es presentar las bases y lineamientos generales del trabajo, que sirven como contexto y guía organizativa de los contenidos contemplados al momento de abordar su lectura.

Capítulo 2 - Aprendizaje por transferencia

2.1 - Introducción

En este capítulo, se establecerá el marco teórico de la tesina. Más precisamente, se introducirán las principales motivaciones que posibilitaron el surgimiento y desarrollo del área; los objetivos concretos que se persiguen con su aplicación; subáreas y paradigmas existentes; conceptos y definiciones fundamentales; contextos aplicativos asociados y factores eventualmente limitantes. Finalmente, se introduce un tópico novedoso en el área, el aprendizaje de larga vida, a fin de comprender las problemáticas que apunta a resolver.

2.2 - Concepto

El **aprendizaje por transferencia (TL)** ha emergido como un área de investigación dentro del **aprendizaje automático (ML)**, que focaliza en la generación e incorporación de conocimiento durante la resolución de un determinado problema y su posterior aplicación sobre otro diferente pero relacionado. [1]

Como se ilustra en la **figura 2.1**, a diferencia del ML tradicional, que genera un sistema (**modelo**) de aprendizaje donde cada tarea se aprende desde cero, el TL construye una **base de conocimiento (KB)** a partir del entrenamiento de tareas (de origen), transferible hacia otra (de destino u objetivo) posibilitando la flexibilización o reducción de requerimientos en cuanto a cantidad y calidad de datos de entrenamiento.

Entonces, mientras que los algoritmos tradicionales de ML y semi-supervisados efectúan predicciones mediante modelos matemáticos entrenados sobre los datos (etiquetados o no) previamente recopilados; el TL admite que tanto dominios como tareas, espacios y distribuciones de características subyacentes, involucrados en fases de entrenamiento y prueba, difieran. De esta forma, prescinde de la recolección y procesamiento de nuevos datos de entrenamiento, reconstrucción y reentrenamiento de modelos desde cero, reduciendo esfuerzos gracias a la transferencia y explotación del conocimiento extraído desde dominios de origen similares, que permitan la construcción de modelos reutilizables sobre otros dominios destino. [4]

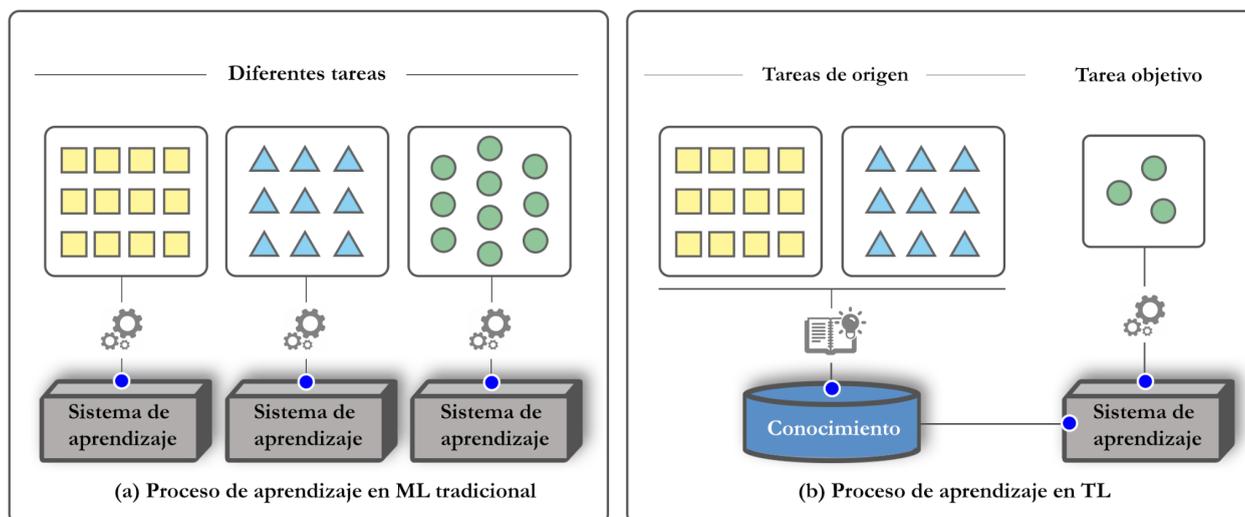


Figura 2.1: Diferencias entre el proceso de aprendizaje en ML tradicional y TL

Figura adaptada de **A Survey on TL - Pan y Yang**

2.3 - Motivación

A pesar de que el **ML** tradicional ha alcanzado un sorprendente éxito y ha sido aplicado mostrando grandes resultados, el escenario ideal que plantea es la abundancia de datos etiquetados que respeten la misma distribución en etapas de entrenamiento y prueba, situación que además de ser trabajosa y costosa en tiempo, resulta poco realista en muchos escenarios actuales del mundo real. Cuando la suposición de que los datos de entrenamiento y pruebas se extraen del mismo espacio de características y conservan idénticas distribuciones de probabilidad subyacentes no se cumple, los modelos necesitan ser reconstruidos con nuevos datos de entrenamiento, resultando en un proceso costoso y en ocasiones impracticable. [1] [2]

El estudio del **TL** se encuentra motivado en la capacidad de las personas para aplicar inteligentemente el conocimiento aprendido para resolver nuevos problemas, reduciendo costo y esfuerzo gracias a la aplicación de soluciones progresivamente más eficientes. Como aprendices, al afrontar tareas novedosas poseemos formas inherentes de transferencia mediante el reconocimiento y aplicación del conocimiento relevante recabado desde experiencias previas de aprendizaje. A mayor relación entre estas experiencias y las tareas novedosas que afrontamos, más simple será su dominio, surgiendo la necesidad de aplicar métodos de ML a largo plazo que permitan la reutilización del conocimiento previo disponible. [4] [28]

Otra motivación surge cuando los datos son susceptibles a quedar desactualizados debido al alto grado de dinamismo (temporal, espacial o dependiente de dispositivos) que ciertas tareas presentan. Por otra parte, en escenarios realistas es extraño contar con abundancia de datos de entrenamiento etiquetados o de dominio público, lo cuál dificulta su recopilación, pudiendo requerir de etapas de preprocesamiento. Surge entonces la idea de adaptar un **modelo** previamente entrenado en la resolución de tareas previas sobre diferentes dominios (origen) para posteriormente transferir el conocimiento o habilidades aprendidas a fin de potenciar el aprendizaje de una nueva **tarea** relacionada y **dominio** diferente (objetivo) valiéndose de datos de entrenamiento estándar y una fuente de información adicional preexistente; como ilustra la **figura 2.2**. [1] [4]

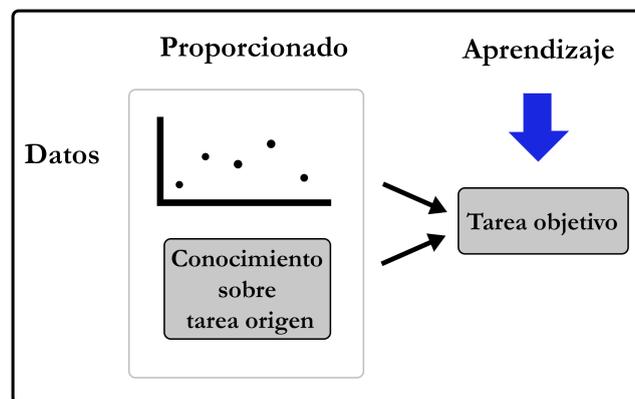


Figura 2.2: Fuentes de información en el TL
Figura adaptada de **Transfer Learning** - Torrey y Shavlik

En el campo del ML tradicional, dado que la disponibilidad de un conjunto de datos con gran número de **instancias** será determinante para el rendimiento alcanzable por el modelo generado, surge la necesidad de acceso a conjuntos de datos cuantitativos y cualitativos (usualmente de dominio público) que se adapten a determinada problemática a resolver y permitan obtener elevados niveles de rendimiento.

Otro factor motivacional para el desarrollo del TL gira en torno al consumo de recursos necesarios para el entrenamiento de complejos modelos de ML sobre enormes conjuntos de datos. El desarrollo de soluciones de TL permitirá la ejecución de modelos sobre plataformas que posean recursos computacionales limitados, modelos pre-entrenados admitirán conjuntos de datos de destino de reducido tamaño y complejidad, así como la utilización de herramientas de software que aprovechen el procesamiento en la nube.

Las etapas propias del proceso generativo de los modelos de ML, como: recopilación, análisis, ingeniería y preprocesamiento de datos, entrenamiento, puesta a punto, validación y prueba, puede requerir tiempos extensos (horas o incluso días) cuando se efectúan desde cero. Una de las motivaciones centrales para el desarrollo y progreso de la técnica de TL se relaciona con la reducción de estos períodos de tiempo.

Finalmente, la implantación de soluciones de ML es un proceso que reviste elevada complejidad y en consecuencia demanda profesionales idóneos y expertos. La aplicación de técnicas de TL también contribuye a relajar este requerimiento, permitiendo que profesionales con un conocimiento comparativamente menor, desplieguen de manera exitosa soluciones de alta complejidad.

Muchas de las motivaciones previamente enunciadas desembocan, directa o indirectamente, en un menor costo de desarrollo e implantación de las soluciones, ya sea en tiempo, esfuerzo, conocimientos requeridos o recursos de hardware. Como motivación ulterior de significativa importancia podemos remarcar que, existe la necesidad de crear sistemas aprendices de alto rendimiento para un dominio objetivo claramente identificado y que son entrenados desde un dominio fuente relacionado pero no necesariamente idéntico. [29]

2.4 - Objetivo

Las técnicas propias del TL tienen por objetivo proporcionar un marco de trabajo para la explotación del conocimiento previamente adquirido gracias a la resolución de problemáticas relacionadas, conocimiento muchas veces consolidado como modelos pre entrenados, a fin de facilitar el modelado predictivo objetivo sobre tareas o dominios destino diferentes. Su principal objetivo reside en la mejora sustancial del aprendizaje de una tarea destino novedosa mediante la transferencia del conocimiento generado durante el proceso de aprendizaje de otra tarea de origen relacionada. En consecuencia, posibilitando la reducción de instancias de entrenamiento etiquetadas que son requeridas en el dominio destino.

Existen distintos factores beneficiosos sobresalientes que la aplicación de técnicas de TL permiten, algunos de los cuales se citan a continuación y pueden visualizarse en la **figura 2.3**, a fin de clarificar las mejoras descriptas:

1. **Rendimiento inicial alcanzable** sobre la tarea destino utilizando el conocimiento transferido y sin otro aprendizaje adicional, comparado al rendimiento inicial de un agente ignorante. Se evidencia cuando los valores de rendimiento iniciales con TL son notablemente superiores.
2. **Tiempo de aprendizaje** completo de la tarea destino dado el conocimiento transferido, comparado con el tiempo que conlleva efectuar el aprendizaje desde cero. Se evidencia en la marcada diferencia entre rendimientos de modelos con o sin TL para cada iteración.
3. **Rendimiento final alcanzable o asintótico** sobre la tarea destino, comparado al alcanzable sin utilizar transferencia de conocimiento. Se evidencia en la proyección de rendimiento futuro alcanzada por los modelos con TL respecto de los que no la utilizan.

4. **Tiempo hasta el umbral**, representa el tiempo de aprendizaje necesario para alcanzar cierto nivel de desempeño preestablecido, que puede verse reducido significativamente utilizando TL. Se evidencia analizando la diferencia temporal o en el número de iteraciones que demoran los modelos con y sin TL para alcanzar un rendimiento prefijado.

[1] [2] [28] [44]

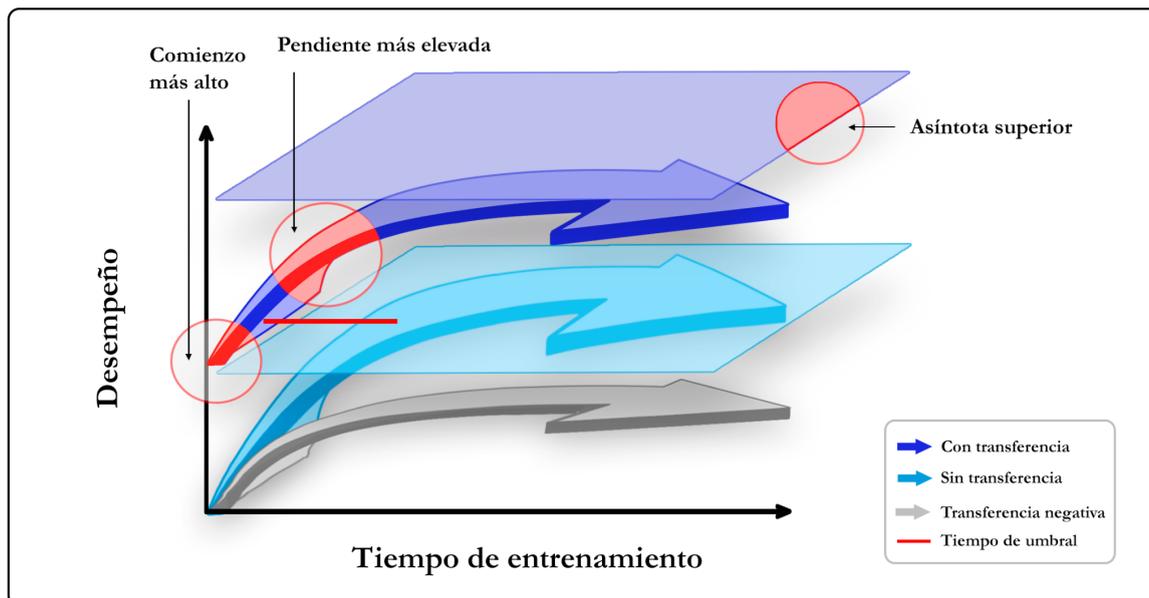


Figura 2.3: Factores de mejora de la transferencia respecto del aprendizaje
 Figura adaptada de **Transfer Learning - Torrey y Shavlik**

Uno de los principales objetivos en el desarrollo y aplicación de métodos de transferencia es que la resolución de tareas innovadoras implique un incremento representativo y ponderable del rendimiento en comparación con la utilización de técnicas tradicionales de ML, escenario conocido como **transferencia positiva (PT)**, y evitando enfáticamente su disminución respecto del aprendizaje sin transferencia, escenario conocido como **transferencia negativa (NT)**. El análisis de factores influyentes para la ocurrencia de estas situaciones es trascendental ya que resume el éxito o fracaso en la consecución de uno de los objetivos principales que persigue la aplicación de técnicas de TL, en la figura previa pueden observarse esos escenarios.

La aplicación de metodologías de TL también impactan positivamente sobre el grado de generalización de los **modelos de ML**, permitiendo la utilización de conjuntos de datos de menor tamaño (con decenas o cientos de **instancias**) y mayor simplicidad (baja **dimensionalidad**), posibilitando el ahorro de tiempo de **aprendizaje** (aceleración) y restringiendo los recursos de cómputo requeridos, relajando la necesidad imperiosa de hardware de alta prestaciones, favoreciendo además el elevado rendimiento de los modelos con un menor número de iteraciones de entrenamiento debido al incremento de eficiencia y desempeño, y generando representaciones más robustas y la flexibilización de restricciones que ocasiona la consideración de un espectro aplicativo más amplio. Como desventajas o factores de consideración, podemos mencionar que ciertas metodologías requieren extremar los cuidados a fin de evitar situaciones indeseadas (como **subajuste** o **sobreajuste**) en los modelos y además, se debe considerar que la aplicación de ciertos algoritmos pueden requerir conocimientos técnicos y experimentales más profundos, propios de expertos.

2.5 - Definiciones

Dominio

Un dominio, que nomenclaremos como D , está compuesto por:

1. Un **espacio de características** X .
2. Una **distribución de probabilidad marginal** $P(X)$ donde $X = \{X_1, \dots, X_n\} \in X$, siendo X un conjunto de instancias, $X = \{x | x_i \in X, i = 1, \dots, n\}$.

En general, se dice que dos dominios $D = \{X, P(X)\}$ difieren, en caso que no coincidan sus espacios de características o sus distribuciones de probabilidad marginal. [1]

Tarea

Dado un **dominio** específico, una tarea, que nomenclaremos como T , está compuesta por:

1. Un **espacio de etiquetas** Y .
2. Una **función predictiva** objetivo $f(\cdot)$ que está implícita, es decir, no se observa pero se aprende a partir de los datos de entrenamiento (pares $\{x_i, y_i\}$, donde $x_i \in X, y_i \in Y$), y puede usarse para predecir la etiqueta $f(x)$ correspondiente a una **instancia** x .

En general, se dice que dos tareas $T = \{Y, f(\cdot)\}$ difieren, en caso que no coincidan sus espacios de etiquetas o sus distribuciones de probabilidad asociadas.

Si bien es posible que en escenarios de TL existan múltiples dominios y tareas de origen así como múltiples dominios y tareas destino, por simplicidad, en este trabajo sólo consideraremos los casos en que existe un único dominio origen (denominado D_S) y un único dominio destino (denominado D_T) como asume la mayoría de los trabajos. Más específicamente, denotamos el dominio origen de datos como $D_S = \{(x_{S1}, y_{S1}), \dots, (x_{Sn}, y_{Sn})\}$, donde $x_{Si} \in X_S$ corresponde a una instancia del conjunto de datos y $y_{Si} \in Y_S$ es la etiqueta de clase asociada a dicha instancia. De manera análoga, definimos el dominio objetivo como $D_T = \{(x_{T1}, y_{T1}), \dots, (x_{Tn}, y_{Tn})\}$, donde la entrada $x_{Ti} \in X_T$ y $y_{Ti} \in Y_T$ es la correspondiente salida. Generalmente, el número de instancias presentes en D_T suele ser significativamente menor al de D_S ($0 \leq n_T \ll n_S$). [1]

Aprendizaje por transferencia (TL)

Dado un **dominio** (D_S) y **tarea** (T_S) de origen y un **dominio** (D_T) y **tarea** (T_T) de destino; el TL optimiza el aprendizaje de una **función predictiva** objetivo $f_T(\cdot)$ desarrollada sobre D_T gracias al aprovechamiento del conocimiento incorporado desde D_S y T_S , donde $D_S \neq D_T$ o bien $T_S \neq T_T$.

Un dominio se define como un par $D = \{X, P(X)\}$, por consiguiente que dos dominios difieran implica que bien los **espacios de características** asociados difieren ($X_S \neq X_T$) o bien que ambos espacios coinciden (se encuentran implícita o explícitamente relacionados) pero las **distribuciones de probabilidad marginal** entre dominios difieren ($P_S(X) \neq P_T(X)$).

De forma análoga a lo expuesto previamente, una tarea se define como un par $T = \{Y, P(Y|X)\}$, entonces que dos tareas de aprendizaje difieran implica que bien los **espacios de etiquetas** asociados a ambos dominios difieren ($Y_S \neq Y_T$) o bien que las **distribuciones de probabilidad condicional** de ambos dominios difieren $P(Y_S|X_S) \neq P(Y_T|X_T)$. [1]

2.6 - Técnicas

La investigación relativa al TL gira en torno a tres tópicos principales: **¿qué transferir?**, dado que algunas secciones del conocimiento son específicas a cada dominio o tarea mientras que otras poseen rasgos compartidos útiles para la mejora del rendimiento mediante transferencia, es importante descubrir y delimitar claramente dichas secciones; **¿cómo transferir?**, identificadas las secciones a transferir, se deben desarrollar algoritmos adecuados para la transferencia efectiva; **¿cuándo transferir?**, debido a que no siempre la aplicación de técnicas de TL resulta beneficiosa (**PT**), pudiendo incluso resultar perjudicial (**NT**) y sumado a que su implementación no representa un proceso sencillo, deben evaluarse todos aquellos factores intervinientes a fin de discernir la viabilidad e impacto de su aplicación.

De acuerdo a la **definición de TL** enunciada y al grado de relación de tareas y dominios, es posible establecer diferentes técnicas, como detalla la **tabla 2.1** donde apreciamos que es en situaciones donde se admite que **dominios** o **tareas**, de origen y destino difieran, cuando el TL emerge como rama resolutive y en cierto sentido, evolutiva del ML tradicional. Sumado a esto, y tomando en cuenta la disponibilidad o ausencia de datos etiquetados, podemos notar que surgen diversas metodologías, como detalla la **tabla 2.2**. [1]

Técnica de aprendizaje		Dominios de origen y destino	Tareas de origen y destino
ML tradicional		Idénticos	Idénticas
TL	Inductiva	Idénticos	Diferentes pero relacionadas
	Transductiva	Diferentes pero relacionados	Idénticas
	No supervisada	Diferentes pero relacionados	Diferentes pero relacionadas
	Por refuerzo	Diferentes pero relacionados	Diferentes pero relacionadas

Tabla 2.1: Relación entre ML tradicional y técnicas de TL

Tabla adaptada de **A Survey on TL - Pan y Yang**

2.6.1 - Transferencia inductiva

Definición: dado un dominio D_S y una tarea T_S de aprendizaje origen; un dominio D_T y una tarea T_T de aprendizaje destino; el **TL inductivo (ITL)** tiene por objetivo perfeccionar el aprendizaje de una **función predictiva** objetivo $f_T(\cdot)$ en D_T utilizando el conocimiento de origen (D_S y T_S), donde $D_S = D_T$ y $T_S \neq T_T$. [1]

Esta técnica requiere la presencia de datos etiquetados en D_T a fin de inducir un **modelo predictivo** objetivo. Además, considerando la situación en la que se disponga de datos etiquetados o únicamente sin etiquetar en D_S podemos subclasificar esta técnica en:

- Inexistencia de datos etiquetados disponibles en D_S , determinando un contexto de aprendizaje similar al **aprendizaje autodidacta (StL)**, donde el espacio de etiquetas entre dominios puede diferir, haciendo que la información parcial en D_S no sea utilizable directamente.
- Existencia de datos etiquetados disponibles en D_S , determinando un contexto de aprendizaje similar al **aprendizaje multitarea (MtL)**, sin embargo, a diferencia de **MtL** que intenta aprender múltiples tareas de origen y destino simultáneamente, aquí el objetivo es un alto rendimiento sobre T_T gracias a la transferencia de conocimiento proveniente desde T_S .

2.6.2 - Transferencia transductiva

Definición: dado un dominio D_S y una tarea T_S , de aprendizaje de origen; un dominio D_T y una tarea T_T , de aprendizaje de destino; el **TL transductivo (TTL)** tiene como objetivo perfeccionar el aprendizaje de la **función predictiva** objetivo $f_T(\cdot)$ en D_T utilizando el conocimiento de origen (D_S y T_S), donde $D_S \neq D_T$ y $T_S = T_T$. [1]

En esta técnica, la coincidencia entre **tareas** permite la adaptación de la **función predictiva** aprendida desde D_S para su utilización sobre D_T incluso cuando éste contenga datos sin etiquetar, mientras que los **dominios** difieren (aunque se encuentran relacionados). Aquí no se dispone de datos etiquetados en D_T pero si suele disponerse de gran cantidad de datos etiquetados en D_S . El término transductivo, en el contexto de transferencia del conocimiento, enfatiza el hecho de que las tareas deben ser coincidentes y la presencia de datos sin etiquetar disponibles en D_T . Además, en base a las definiciones de **TL transductivo** y **dominio** previamente citadas, sabemos que ambos dominios difieren, por consiguiente se contemplan dos subclasificaciones diferentes en esta técnica dependiendo si los dominios difieren en los **espacios de características** o en la **distribución de probabilidad marginal** entre ellos:

- El espacio de características entre dominios de origen y destino involucrados (D_S y D_T) difieren ($X_S \neq X_T$).
- El espacio de características entre dominios de origen y destino involucrados (D_S y D_T) coinciden ($X_S = X_T$), pero las distribuciones de probabilidad marginal de los datos de entrada difieren ($P(X_S) \neq P(X_T)$).

El caso inicial es considerado como transferencia de dominios heterogéneos (**HetTL**) y se suelen utilizar adaptaciones como el **sesgo de selección de muestra (SSB)** o el **cambio covariante (CS)**; por su parte, el último caso es considerado como transferencia de dominios homogéneos (**HomTL**) y se relaciona con la **adaptación de dominios (DA)**.

2.6.3 - Transferencia no supervisada

Definición: dado un dominio D_S y una tarea T_S de aprendizaje de origen; un dominio D_T y una tarea T_T de aprendizaje destino; el **TL no supervisado (UTL)** tiene como objetivo perfeccionar el aprendizaje de una **función predictiva** objetivo $f_T(\cdot)$ sobre D_T utilizando el conocimiento previamente incorporado desde el origen (D_S y T_S) donde $D_S \neq D_T$ y $T_S \neq T_T$, siendo además Y_S y Y_T no observables. [1]

En esta técnica, tanto dominios como tareas involucradas en el proceso de transferencia difieren (aunque se encuentran relacionados) y la inobservancia de Y_S y Y_T , se refiere a que no se dispone de datos etiquetados en ninguno de los dominios durante la fase de entrenamiento. Concentrándose en la resolución de tareas no supervisada sobre D_T tal como lo hacen las metodologías de **agrupamiento, reducción de dimensionalidad** y **estimación de densidad**.

Técnica de TL	Áreas relacionadas	Datos etiquetados en el dominio de origen	Datos etiquetados en el dominio de destino	Metodologías
Inductiva	Aprendizaje multitarea	Disponible	Disponible	Clasificación y regresión
	Aprendizaje autodidacta	No disponible	Disponible	Clasificación y regresión
Transductiva	Adaptación de dominios Sesgo de selección de muestra Cambio covariante	Disponible	No disponible	Clasificación y regresión
No supervisada		No disponible	No disponible	Agrupamiento Reducción de dimensionalidad Estimación de densidad

Tabla 2.2: Técnicas de TL y disponibilidad de datos etiquetados en dominios
Tabla adaptada de **A Survey on TL - Pan y Yang**

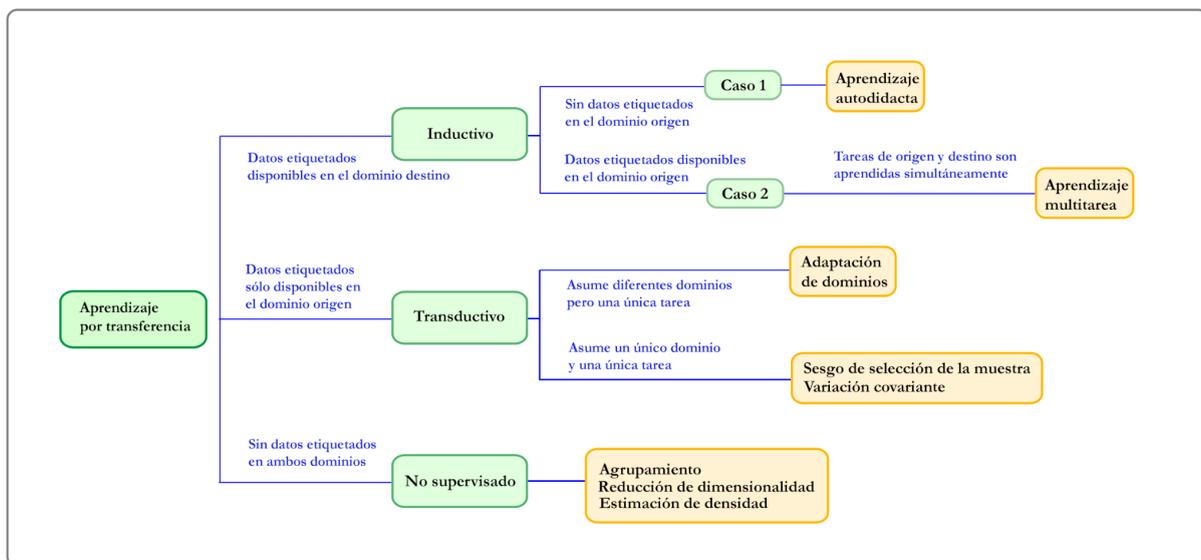


Figura 2.4: Diagrama integral de técnicas y metodologías de TL
Figura adaptada de **A Survey on TL - Pan y Yang**

2.6.4 - Transferencia por refuerzo

En esta técnica de TL, también denominada **aprendizaje por transferencia reforzado (RTL)** un **agente de aprendizaje** trabaja en un entorno de control secuencial denominado **proceso de decisión de Markov (MDP)**, monitoreando el estado de su entorno y efectuando acciones que lo alteran y desencadenan recompensas. Su objetivo es la resolución de un problema temporal de asignación de créditos, mediante el aprendizaje de una política de acción que maximice la recompensa acumulada.

Un agente **RL** típico se comporta de acuerdo al diagrama mostrado en la **figura 2.5**, donde en el paso temporal t , sondea su estado (e_t) y política actual (π) para realizar una acción ($\pi(e_t) = a_t$). Luego, recibe una recompensa (r_t) y avanza a un nuevo estado temporal (e_{t+1}). A partir de la utilización de esta información actualiza su política antes de repetir la operatoria de forma iterativa. A menudo RL consiste de una secuencia de episodios que finalizan cuando el agente alcanza un estado final. Además, conforme avanza el proceso de aprendizaje, el agente debe equilibrar entre explotar la política actual que lo alienta a operar en áreas de altas recompensas o explorar otras nuevas, potencialmente superadoras.

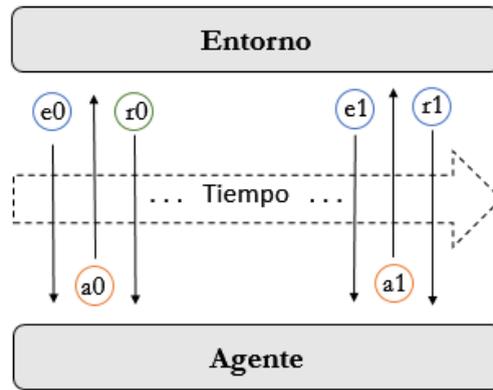


Figura 2.5: Interacción de un agente de aprendizaje por refuerzo con su entorno
 Figura adaptada de **Transfer Learning - Torrey y Shavlik**

2.7 - Enfoques

Considerando la interrogante previamente planteada “¿qué transferir?” y las técnicas analizadas (inductiva, transductiva, no supervisada y por refuerzo) surgen cuatro enfoques, cuyo tratamiento nos permitirá evaluar los aspectos particulares de cada problemática, a fin de seleccionar y aplicar las metodologías más adecuadas. Los enfoques referidos son los siguientes:

1. **Transferencia de instancias**, donde se considera que ciertas partes de los datos del dominio de origen son útiles en el entrenamiento sobre el dominio objetivo y pueden reutilizarse, mediante la aplicación de procesos previos para determinar su importancia en el entrenamiento. Aquí la re-ponderación de instancias y el muestreo de importancia son técnicas principales.
2. **Transferencia de representación de características**, donde la idea subyacente es el descubrimiento y aprendizaje de una representación satisfactoria de características del dominio destino a fin de reducir diferencias entre dominios y el error del modelo generado. El conocimiento utilizado para transferir datos entre dominios es codificado en una representación de características aprendida, para lograr una mejora sustancial en el rendimiento.
3. **Transferencia de parámetros**, asume que las tareas origen y destino comparten parámetros o distribuciones previas de hiperparámetros de los modelos. El conocimiento transferido es codificado en parámetros compartidos descubiertos o previos entre modelos, resultando beneficioso para la transferencia.
4. **Transferencia de conocimiento relacional**, supone la existencia de relaciones entre los datos de ambos dominios, asumiendo que se trata con dominios relacionales y relajando la restricción de ser **i.i.d.** Por lo tanto, esas relaciones entre dominios representan el conocimiento a transferir mediante la construcción de mapeos de conocimiento relacional.

[1]

En términos relativos, la técnica inductiva ha sido ampliamente investigada bajo cada enfoque descrito, mientras que la técnica transductiva ha experimentado un grado de investigación intermedio, y ha sido examinada bajo los enfoques de transferencia de instancias y representación de características, por su parte la técnica de transferencia no supervisada representa un tópico de investigación más novedoso y relativamente menos abordado, únicamente estudiado bajo el enfoque de representación de características; ver **tabla 2.3**.

Técnica de TL \ Enfoque	Transferencia de instancias	Transferencia de representación de características	Transferencia de parámetros	Transferencia de conocimiento relacional
Inductiva	Sí	Sí	Sí	Sí
Transductiva	Sí	Sí	No	No
No supervisada	No	Sí	No	No

Tabla 2.3: Técnicas de TL y enfoques utilizados
 Tabla adaptada de **A Survey on TL - Pan y Yang**

2.7.1 - Transferencia de instancias

Aprendizaje por transferencia inductiva

En el enfoque de transferencia de instancias para la técnica de transferencia inductiva (ITL), los datos del dominio de origen no pueden ser utilizados directamente, pero sí existen ciertas partes reutilizables de manera conjunta con datos etiquetados disponibles en el dominio destino.

Existen algoritmos (como **TrAdaBoost**, extensión de **AdaBoost**) que asumen dominios con idénticos conjuntos de características y etiquetas pero distribuciones de probabilidad diferentes, resultando algunos datos del dominio de origen útiles para el aprendizaje pero otros inútiles o incluso dañinos; entonces se propone un proceso iterativo de reponderación, a fin de alentar la utilización de aquellos datos “adecuados” minimizando el efecto adverso resultante de la utilización de los “inadecuados”. Por otro lado, se han propuesto métodos heurísticos para la eliminación de ejemplos de entrenamiento “engañosos” del dominio de origen basados en la diferencia de probabilidades condicionales de etiquetas respecto de características entre dominios ($P(Y_T | X_T)$ y $P(Y_S | X_S)$). También existen métodos de aprendizaje activo para la selección de datos no etiquetados en el dominio destino y etiquetarlos con ayuda de los datos del dominio de origen.

Aprendizaje por transferencia transductiva

La mayoría de los enfoques bajo esta técnica están motivados por métodos de **muestreo de importancia (IS)** donde previamente se analiza y resuelve el problema de **minimización del riesgo empírico (ERM)**, a fin de aprender la parametrización óptima de los modelos que minimice el riesgo esperado ya que no se dispone de datos etiquetados en D_T , se debe aprender un modelo desde D_S resolviendo diferentes problemas dependiendo de la relación entre distribuciones de probabilidad de ambos dominios. Si ambas distribuciones de probabilidad en D_S coinciden, podemos aprender el modelo mediante la resolución de un problema de optimización y utilizarlo sobre D_T , caso contrario se debe aprender un modelo con elevada capacidad de generalización y precisión mediante la incorporación de valores de penalización de instancias con sus correspondientes pesos y en casos donde la diferencia de distribuciones de probabilidad entre dominios sea causada por las diferentes distribuciones de probabilidad marginal del espacio de características de ambos dominios, mediante la estimación de la relación de probabilidades entre instancias.

2.7.2 - Transferencia de representación de características

Aprendizaje por transferencia inductiva

Este enfoque apunta al descubrimiento de “buenas” representaciones de características en D_T , con el objetivo de minimizar la divergencia entre los dominios involucrados en el proceso de transferencia, así como del error del modelo resultante. Las estrategias para construir dichas representaciones dependen de la presencia o ausencia de datos etiquetados en el dominio de origen:

- Si **existen datos etiquetados en D_S** , se utilizan métodos de aprendizaje supervisado (similar al aprendizaje de características comunes en **MtL**) intentando aprender una representación de baja dimensionalidad compartida entre tareas diferentes pero relacionadas.
- Si **no existen datos etiquetados en D_S** , se utilizan métodos de aprendizaje no supervisado para el aprendizaje de características de alto nivel (similar al **algoritmo de codificación dispersa**) recientemente adaptados para TL.

Aprendizaje por transferencia transductiva

La mayoría de las metodologías se ubican bajo entornos de aprendizaje no supervisados. Se han desarrollado algoritmos **SCL** a fin de utilizar los datos sin etiquetar en D_T para la extracción de características relevantes que minimicen la diferencia entre dominios, se han propuesto métodos heurísticos para el etiquetado de oraciones en entornos **NLP** y el desarrollo posterior del **método de información mutua (MIM)**, así como una combinación de estos (**SCL-MI**) que busca características pivot con gran dependencia de las etiquetas en D_S . En el dominio de **NLP**, referido como **adaptación de dominio**, se han propuesto métodos que utilizan una función de correspondencia nuclear entre dominios; pero su dificultad de aplicación sobre otras áreas motivó la investigación de los límites de la convergencia uniforme para algoritmos de minimización de combinación convexa del riesgo empírico de origen y destino.

Aprendizaje por transferencia no supervisada

Este enfoque considera un nuevo tipo de problemas de agrupamiento, conocido como **agrupamiento autodidacta (StC)**, cuyo objetivo es agrupar una pequeña cantidad de datos sin etiquetar en D_T con la ayuda de gran cantidad de datos, también sin etiquetar en D_S . Aprendiendo un espacio de características comunes entre dominios que ayuda al agrupamiento en D_T . También se propuso un algoritmo de **análisis discriminativo transferido (TDA)** para resolver el problema de reducción de dimensionalidad de transferencia intentando hallar el mejor subespacio mediante la aplicación iterativa y alternada de métodos de agrupamiento para generar etiquetas de pseudo-clases para datos no etiquetados de D_T seguido de métodos de reducción de dimensionalidad de los datos destino y los etiquetados de origen.

2.7.3 - Transferencia de parámetros

Aprendizaje por transferencia inductiva

La mayoría de las metodologías de este enfoque suponen que los modelos individuales de tareas relacionadas deberían compartir ciertos parámetros o distribuciones previas de hiperparámetros. La mayoría de los algoritmos incluyen un entorno de regularización y bayesiano jerárquico diseñados para trabajar bajo **MtL** adaptables al **TL**. En MtL los pesos de las funciones de pérdida de los datos de origen y destino son los mismos, mientras que en TL éstos pueden variar ya que intuitivamente, podemos asignar un mayor peso a la función de pérdida de D_T para alcanzar un mejor rendimiento.

Se ha propuesto la utilización de **IVMs** multitarea (**MT-IVMs**) que intentan aprender parámetros de un **proceso Gaussiano (GP)** sobre múltiples tareas que comparten el mismo GP. Además se ha propuesto el uso de una **matriz de covarianza** sobre las tareas para modelar sus dependencias mediante un GP previo usado para inducir las correlaciones. También se ha propuesto el uso de un entorno bayesiano jerárquico conjuntamente con GP; mientras que otros investigadores han propuesto transferir parámetros de **SVMs** bajo un entorno de regularización. Finalmente, se ha propuesto un **marco de trabajo de aprendizaje de conjunto ponderado localmente (LWE)** para combinar múltiples modelos para TL, donde los pesos son asignados dinámicamente de acuerdo con el poder predictivo del modelo ante cada ejemplo de prueba en D_T .

2.7.4 - Transferencia de conocimiento relacional

Aprendizaje por transferencia inductiva

A diferencia de los restantes enfoques de transferencia bajo la técnica inductiva, este trata con problemas de transferencia sobre dominios relacionales donde los datos no necesariamente son considerados **i.i.d.**, como se supone tradicionalmente, y además pueden representarse mediante múltiples relaciones. El objetivo es intentar transferir la relación existente entre los datos de ambos dominios y las técnicas estadísticas de aprendizaje relacional son las principales bajo este enfoque.

El algoritmo **TAMaR** transfiere el conocimiento entre dominios relacionales haciendo uso de **MLNs** basándose en la existencia y descubrimiento de mapeos relacionales entre dominios (ver **sección 3.6** para más detalles). También se han propuesto metodologías de transferencia del conocimiento relacional basadas en la lógica de segundo orden para el descubrimiento de regularidades estructurales en D_S mediante fórmulas lógicas de Markov y su posterior instanciación con variables predicado pertenecientes a D_T .

2.8 - Limitaciones

Un tópico que reviste notable importancia es el que comprende el análisis y reconocimiento de los límites que la técnica de TL trae aparejados, ya que existen circunstancias en las cuales su aplicación no es recomendable ante el riesgo de deterioro del rendimiento (**NT, transferencia negativa**), objetivo contrapuesto al perseguido (**PT, transferencia positiva**). Por ende, el estudio de las metodologías adecuadas para evitar que NT acontezca, representa un punto crucial de la investigación del área. [1]

La efectividad del TL depende en gran medida del grado de relación entre las tareas involucradas. Si ambas se encuentran estrechamente relacionadas, el método de TL aplicado puede aprovechar esa situación, produciendo un incremento del rendimiento sobre el aprendizaje de la tarea objetivo. Por el contrario, si el grado de relación entre tareas no es fuerte o la relación no se encuentra adecuadamente afianzada por el algoritmo de TL aplicado, el rendimiento obtenido mediante transferencia puede experimentar una muy reducida mejoría, no experimentar mejoría alguna, o incluso sufrir una marcada disminución. [2]

Idealmente, desearíamos que un método de TL tenga la suficiente flexibilidad para producir **PT** al trabajar con tareas fuertemente relacionadas y evite **NT** con aquellas débilmente relacionadas. En la práctica, este doble objetivo es difícil de alcanzar dado que los enfoques que ofrecen garantías contra NT a menudo limitan el efecto de PT y por su parte, los enfoques más agresivos potencian los efectos de la PT pero, a menudo, no tienen resguardo contra la NT; la **figura 2.6** grafica el escenario descripto. [2]

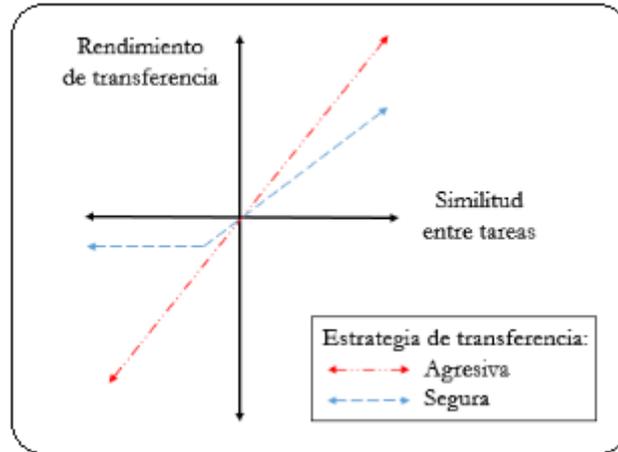


Figura 2.6: Métodos de TL y su relación con el rendimiento y grado de afinidad entre tareas
 Figura adaptada de **Transfer Learning - Torrey y Shavlik**

Para evitar la ocurrencia de NT, debe estudiarse el grado de transferibilidad que presentan los dominios y tareas involucrados, estableciendo métricas de similitud y grado de afectación de transferibilidad, a fin que modelo generado mediante TL no evidencie un rendimiento inferior al que se hubiese obtenido aprendiendo la tarea objetivo desde cero. Entre los principales métodos para enfrentar esta situación podemos mencionar:

- **Rechazo de información incorrecta**, donde se reconoce y rechaza el conocimiento dañino en T_S mientras se aprende T_T , con el objetivo de minimizar el impacto negativo que la incorporación de información incorrecta tiene sobre el rendimiento, equilibrando el alcance de PT y la prevención de NT.
- **Selección de tarea de origen**, donde se asume que la probabilidad de evitar NT aumenta si en lugar de tener una única T_S desde la cual transferir, contamos con un conjunto candidato de ellas. Entonces el problema se transforma en la elección de la T_S más adecuada a partir de ese subconjunto.
- **Modelado de similitud entre tareas**, evalúa los beneficios de utilizar varias o todas las T_S candidatas en lugar de limitarse a seleccionar sólo una de ellas, como hacía el método previo.

[2]

2.9 - Aprendizaje automático de largo plazo

Existe un novedoso tipo de **ML** denominado aprendizaje automático de largo plazo, para toda la vida o permanente (**LML**). El paradigma tradicional de ML es considerado aislado, en el sentido que no considera información relacionada aprendida y al carecer de memoria, no retiene el conocimiento incorporado ni se lo utiliza para futuros aprendizajes. Tal característica estructural genera que estos sistemas requieran abundancia de ejemplos de entrenamiento para concretar un aprendizaje efectivo, situación que se ve agravada en paradigmas como el supervisado, donde el proceso de etiquetado de datos resulta costoso, o ante tareas con alto grado de dinamismo que exigen un re-etiquetado continuo. [6]

Definición: LML es un proceso de aprendizaje continuo, donde un agente aprendiz que ha concretado una secuencia de N tareas de aprendizaje, T_1, T_2, \dots, T_n ; al abordar la tarea T_{n+1} con sus respectivos datos D_{n+1} , puede aprovechar el conocimiento acumulado en su **base de conocimiento (KB)** durante el

aprendizaje de las n tareas previas, para optimizar el aprendizaje de la tarea T_{n+1} . Luego del aprendizaje de T_{n+1} , KB se actualiza con resultados (intermedios y finales) aprendidos desde T_{n+1} . [6]

Los sistemas LML poseen características distintivas respecto de otros métodos de **TL** y **MtL**. A diferencia de los métodos unidireccionales de TL, que utilizan información de origen para ayudar en el destino, los sistemas LML pueden moverse libremente asumiendo dominios involucrados similares, además mientras que TL usa datos etiquetados para el aprendizaje, LML aplica la retención del conocimiento. Por su parte, MtL tiene por objetivo efectuar optimizaciones conjuntas sobre múltiples tareas similares para poder compartir ese conocimiento y alcanzar resultados superadores; sin embargo no se acumula conocimiento a lo largo del tiempo (exceptuando datos), ni existe el concepto de aprendizaje continuo. El LML enfatiza la utilización de un enfoque sistemático bajo el cuál se despliega la retención del conocimiento aprendido y su transferencia selectiva a fin de desarrollar políticas resolutivas más precisas. [3][6]

Entre las características más sobresalientes de LML, consideramos:

- **Aprendizaje efectivo y eficiente**, produciendo ante cada tarea, hipótesis superadoras en rendimiento de generalización respecto de aquellas desarrolladas estrictamente desde ejemplos de entrenamiento; transferir el conocimiento desde KB mediante una cuidadosa selección; sesgando favorablemente el aprendizaje de tareas novedosas y el desarrollo de modelos superadores.
- **Retención efectiva y eficiente**, la retención del conocimiento a largo plazo no debería causar la pérdida del conocimiento asociado a tareas previas ni de la novedosa objetivo, sino incrementar la precisión de tareas previas cuando la tarea retenida se encuentre relacionada. Además, el enfoque de representación del conocimiento debería permitir la selección efectiva del conocimiento desde tareas previas para la transferencia inductiva en el aprendizaje a corto plazo.

Los sistemas deberían resistir la introducción y acumulación de conocimiento erróneo, reteniendo en KB hipótesis de alto grado de generalización que no reduzcan la precisión de las retenidas previamente. De hecho, la integración del conocimiento de nuevas tareas debería incrementar la precisión del conocimiento previo relacionado, brindando métodos eficientes de retención en términos de tiempo y espacio requeridos.

2.10 - Conclusión

En este capítulo, se abordó el concepto de aprendizaje por transferencia (**TL**), así como las motivaciones que potenciaron su surgimiento y sostienen su acelerado desarrollo, además de los principales objetivos que la técnica analizada persigue. Se brindaron definiciones formales cuyo conocimiento resultará esencial para abordar el resto de la tesina y ante cualquier profundización bibliográfica que se desee realizar.

Se presentaron las diferentes técnicas y enfoques más sobresalientes en el área, además de las limitaciones que eventualmente pueden surgir en la aplicación del TL bajo los conceptos de transferencia positiva y negativa. También abordó el novedoso tópico del **LML**, que resulta crucial dado que representa uno de los más importantes pasos evolutivos de los sistemas de transferencia del conocimiento.

El objetivo de este capítulo es brindar el marco teórico de la tesina, que lo confiere como pilar de la misma, adquiriendo su lectura carácter de imprescindible dentro de este trabajo ya que esgrime los fundamentos necesarios para la comprensión y profundización de los temas que se abordarán en capítulos subsiguientes.

Capítulo 3 - Estado del arte

3.1 - Introducción

En este capítulo, se presentará la situación actual de investigación y desarrollo en el campo del **TL**. Dada la amplitud del área y el constante crecimiento en las subáreas constituyentes, se focalizará en el repaso puntual de los trabajos más relevantes hasta el momento. No representa el objetivo principal del capítulo abordar en profundidad cada desarrollo, sino exponer los lineamientos principales y objetivos generales, con el objetivo de facilitar la extracción de los conceptos más relevantes y la problemática que cada trabajo abordado resuelve. Se insta al lector interesado a profundizar sobre cada uno de ellos, a partir de las respectivas referencias bibliográficas citadas.

3.2 - Escenarios motivacionales para TL

Si bien muchas tecnologías de **ML** y **minería de datos** han alcanzado un éxito significativo, la mayoría de estas metodologías se desempeñan satisfactoriamente sólo bajo la suposición de que los datos de entrenamiento y pruebas comparten el mismo **espacio de características** y **distribución de probabilidad** asociada, pero al no cumplirse estas suposiciones en muchos escenarios del mundo actual, los modelos deben ser completamente reconstruídos en un proceso que acarrea elevados costos. En esta sección se presentarán, a modo introductorio, tres escenarios realistas con marcados factores de dinamismo y variabilidad, que evidencian los desafíos que el TL es llamado a resolver, sirviendo además estos factores como impulsores de su acelerado desarrollo y evolución.

Se sugiere que la lectura de la actual sección sea complementada mediante la lectura del **anexo C**, que aborda con un mayor grado de detalle los distintos escenarios de aplicación y desarrollos existentes del área.

3.2.1 - Clasificación de páginas web sin ejemplos negativos

Hwanjo Yu, Jiawei Han, y Kevin Chen-Chuan Chang [7]

La clasificación de páginas web en diferentes categorías de interés representa un paso esencial de la minería web que requiere un laborioso proceso de recopilación y preprocesamiento de instancias de entrenamiento tanto positivos como negativos, siendo éste último tipo particularmente costoso y de recolección manual susceptible al sesgo, que pudiera afectar la precisión de los clasificadores. A fin de encauzar estos inconvenientes propios del subcampo del aprendizaje o clasificación de clase única; se ha desarrollado un marco de trabajo denominado **aprendizaje basado en ejemplos positivos (PEBL)**, que elimina la necesidad de recolección manual de ejemplos de entrenamiento negativos, mediante la aplicación de un **algoritmo de mapeo-convergencia (MCa)** que trabaja en dos etapas: una etapa de mapeo donde se utiliza un clasificador débil que delinea una aproximación inicial de datos con “fuerte” presunción negativa y una etapa de convergencia que ejecuta iterativamente un clasificador interno (como un **SVM**) que maximiza los márgenes, posibilitando el perfeccionamiento progresivo de la aproximación de datos negativos.

Uno de los principales inconvenientes con los clasificadores binarios tradicionales es que requieren un proceso laborioso de recolección de ejemplos tanto positivos como negativos para el entrenamiento. Siendo este último tipo especialmente arduo, debiendo abarcar uniformemente el conjunto universal excluyendo la clase positiva siendo el conjunto de clase positiva significativamente menor que su complemento.

El algoritmo propuesto **MCa** delinea la aproximación de datos potencialmente negativos a partir del análisis de datos positivos y sin etiquetar, gracias a la identificación de características potencialmente positivas, analizando la frecuencia de aparición de las mismas, considerando instancias carentes de características positivas como potencialmente negativas. De esa forma, durante la etapa de mapeo se pueden extraer datos potencialmente negativos a partir de datos sin etiquetar, eliminando el problemático caso de recolección de instancias negativas previamente mencionado. Demostrando empíricamente que **MCa** (con datos positivos y sin etiquetar) supera la precisión de **OSVMs**, resultando tan preciso como un **SVM** tradicional (con datos positivos y negativos etiquetados).

Si bien este trabajo no está relacionado a la utilización de TL, se lo pretende utilizar como ejemplo introductorio a fin de apreciar el proceso de detección de una dificultad asociada a la problemática a resolver y su abordaje mediante el desarrollo de un algoritmo que permite sortear ese problema. En ese aspecto podemos notar como se avizora un punto fundamental de contacto con la idea del TL, al recoger una dificultad, en este caso asociada al proceso de recolección y etiquetado manual de ejemplos negativos, para solventarla gracias al aprovechamiento de datos en abundancia (ejemplos sin etiquetar) para lograr el objetivo deseado. Enfatizando además, en evitar suposiciones restrictivas y poco realistas, admitiendo un mayor grado de generalidad resolutive e incrementando, en consecuencia, la aplicabilidad del método propuesto.

3.2.2 - TL para localización en interiores basada en señales WiFi

Sinno Jialin Pan, Vincent Wechen Zheng, Quiang Yang y Derek Hao Hu [8]

El **problema de localización en interiores basado en señales WiFi (WILP)**, tiene por objetivo la detección de la ubicación de un dispositivo cliente a partir de las señales recibidas desde varios **puntos de acceso (APs)**, siendo considerado de alta complejidad y relevancia por muchas aplicaciones de **IA**, computación ubicua y TL dado que la recolección de datos WiFi reviste un alto grado de dinamismo y dependencia del entorno donde alteraciones contextuales en las señales (ruido, desvanecimiento, múltiples rutas, sombras en entornos cerrados) resultan casi constantes; por otra parte, la distribución de datos se encuentra en constante cambio y dependiente de diversos factores (movimiento humano, cambios de temperatura y humedad). Este contexto es terreno fértil para la transferencia de conocimiento de modelos de localización a través de dimensiones temporales, espaciales y de dispositivos.

Aplicar enfoques tradicionales de **ML** sobre este tipo de problemas resulta costoso debido al proceso de recolección y etiquetado de datos de entrenamiento en forma de pares (ubicación, **RSS**) en un edificio de gran escala, que implica recorrerlo recolectando valores RSS y marcando las coordenadas. Además, los modelos generados fallan al utilizarlos con datos recolectados en diferentes períodos de tiempo, locaciones o dispositivos; debiendo repetir el proceso cada vez que la distribución de la señal sufre alteraciones.

Muchos modelos previos de localización asumen que alguna de las dimensiones de dinamismo de datos en la localización (temporal, espacial o de dispositivo) permanece estática, sin requerir adaptación alguna. Pero esta simplificación dista mucho del problema **WILP** realista donde tales adaptaciones resultan inherentemente necesarias ya que los datos recolectados en fase de entrenamiento difieren significativamente respecto de los de prueba. Por lo tanto, la investigación relativa a la reducción del esfuerzo de adaptación sobre modelos de localización en cada dimensión de variabilidad de datos, representan tópicos principales.

Este trabajo divide el enfoque de transferencia de modelos de localización para afrontar **WILP** en tres categorías, según la dimensión temporal, espacial y de dispositivo. En cuanto a la dimensión temporal, el objetivo es construir un modelo preciso para un período de tiempo a partir de datos WiFi recolectados en un determinado lapso de tiempo. En cuanto a la dimensión espacial, el objetivo es descubrir automáticamente estructuras o conocimiento compartido (**KB**) en el dominio de localización previo, para luego incorporarlo a los datos WiFi obtenidos en diferentes áreas y construir un modelo de localización preciso para el espacio completo. En cuanto a la dimensión de dispositivos, nos motiva a modelar el problema como un problema de aprendizaje multitarea donde se espera que las distribuciones de datos sean muy disímiles entre dispositivos.

El análisis de este trabajo evidencia de qué forma las soluciones basadas en **TL** emergen cuando los datos son propensos a quedar desactualizados con facilidad, donde las distribuciones de los mismos pueden variar significativamente debido al alto factor de dinamismo que acarrea el problema, debido a la variabilidad de las señales WiFi a través de diferentes dimensiones. Surge entonces la necesidad de adaptar modelos previamente entrenados hacia un dominio novedoso. Además se intentan desarrollar métodos de transferencia generales para varios problemas realistas, evitando la imposición de restricciones ficticias que acrecienten la distancia entre el problema objetivo, del que se presenta concretamente en el mundo real.

3.2.3 - Adaptación de dominios para la clasificación de sentimientos

John Blitzer, Mark Dredze y Fernando Pereira [9]

Este trabajo se centra en el **problema de adaptación de dominios (DAP)** para la clasificación de sentimientos a partir de reseñas de productos y entre sus objetivos primordiales se encuentra la extensión del **algoritmo de aprendizaje por correspondencia estructural (SCL)** para reducir el error relativo del proceso de adaptación, así como la obtención del criterio de similitud entre dominios (denominado A-distancia) para la selección de aquellos más adecuados. Este problema contempla una variedad de dominios de aplicación que exigen a que investigadores e ingenieros deban abocarse a coleccionar, cotejar y etiquetar datos ante el surgimiento de cada nuevo dominio, tornándose tales esfuerzos prohibitivos.

A partir de la presencia de datos etiquetados en el **dominio** origen y no etiquetados en ambos dominios, el algoritmo **SCL**, primero selecciona un conjunto de características de pivote con mayor frecuencia de ocurrencia conjunta; luego modela la correlación entre cada característica de pivote y las restantes, entrenando predictores lineales de pivotes para predecir la ocurrencia de cada pivote en los datos no etiquetados de ambos dominios. Cada predictor de pivote tiene asociado un vector de ponderación que determina la correlación entre características no pivotes y su correspondiente pivote. Entonces la adecuada elección de características de pivote permitirá a los predictores discriminar entre reseñas positivas y negativas en ambos dominios; la eficacia del método depende en gran medida, de la elección de características pivote.

Se experimentó con un clasificador entrenado y probado sobre cuatro dominios diferentes (libros, DVDs, artículos electrónicos y aparatos de cocina), existiendo afinidad relacional entre libros y DVDs, así como entre electrónicos y aparatos de cocina, pero siendo estos dos grupos mutuamente disímiles. Se definió un clasificador lineal entrenado sin adaptación como línea base, mientras que se tomó como estándar de oro un clasificador entrenado y probado sobre el mismo dominio; la extensión del método propuesto se definió como **SCL-MI**. Naturalmente, la adaptación entre clasificadores de dominios afines es más sencilla, como se aprecia en la **figura 3.1** donde para cada dominio destino se observan tres grupos de barras propios de los métodos de cada dominio origen; las líneas gruesas hacen referencia a la precisión del estándar de oro.

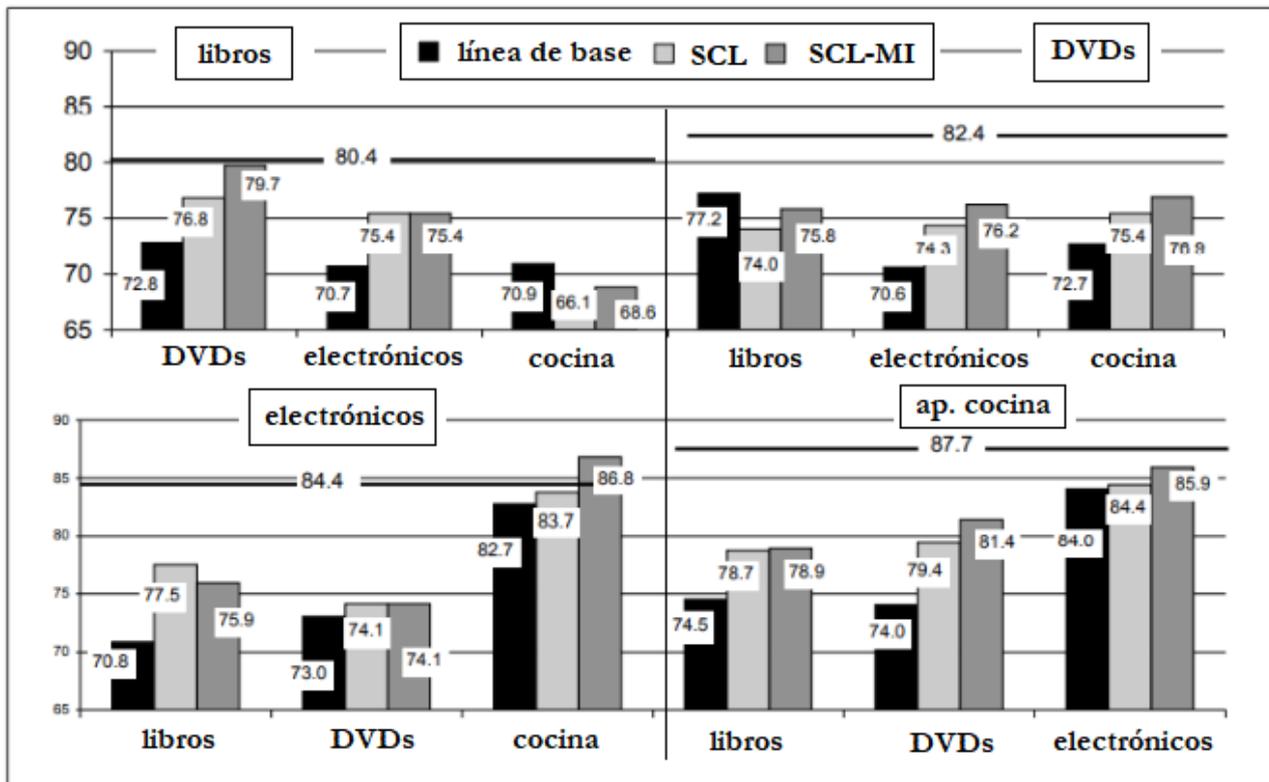


Figura 3.1: Precisión de métodos para la adaptación de dominios
 Figura adaptada de **Domain Adaptation for Sentiment Classification** - Blitzer, Dredze y Pereira

La selección de ciertas características pueden cambiar el comportamiento entre dominios. En ocasiones las características predictivas del dominio origen no son predictivas del dominio destino y otras veces las características predictivas del dominio destino no aparecen en el de origen; SCL trata estos desajustes alineando características entre dominios, mientras que SCL-MI puede utilizar un pequeño número de instancias etiquetadas del dominio destino para aprender a ignorar proyecciones desalineadas, a fin de generar un modelo ajustado con una reducción del error cercana al 46 % respecto del modelo base.

Otro punto sobresaliente de este trabajo es la presentación del concepto de **A-distancia** (distancia de adaptación entre dominios) como métrica de pérdida debida a la adaptación entre dominios, diseñada para tener en cuenta las divergencias que afectan la precisión de clasificación. La misma resulta de gran utilidad, permitiendo seleccionar y etiquetar los dominios de origen más prósperos, para lograr un mayor rendimiento.

Este trabajo aborda dos cuestiones relevantes de la adaptación de dominios, valiéndose de las mejoras de clasificación sentimental alcanzadas por SCL, mediante la selección de características pivote utilizando no sólo la frecuencia común entre dominios sino también la información mutua con las etiquetas de origen y la corrección de sus desajustes aprovechando una pequeña cantidad de datos etiquetados sobre el dominio objetivo (D_t). Además, se brinda un método no supervisado (A-distancia) para la selección del dominio origen (D_s) con el mayor grado de adaptabilidad respecto de D_t , midiendo la divergencia entre dominios y su pérdida de adaptación; posibilitando la selección de aquel D_s que minimice el error de adaptación hacia D_t . Estos factores resaltan el valor fundamental del análisis del grado de adaptabilidad entre dominios, para la obtención de **modelos de ML** más eficientes, reduciendo costos asociados al esfuerzo de etiquetado y preprocesamiento de datos; objetivos que comparte con la aplicación de técnicas de TL.

3.3 - TrAdaBoost: un impulso para TL

Wenyuan Dai, Qiang Yang, Gui-Rong Xue y Yong Yu [19]

Este trabajo se enmarca dentro de la técnica de **TL inductivo (ITL)** dado que algunos datos etiquetados en el D_T se utilizan para inducir y mejorar el aprendizaje de una función predictiva objetivo $f_T(\cdot)$ a partir del conocimiento adquirido desde uno o más dominios y tareas de origen, cuyos datos son utilizados auxiliarmente. Además se encuentra bajo el enfoque de **transferencia de instancias**, ya que si bien no es posible utilizar directamente todos los datos de D_s , sí es factible utilizar ciertas secciones de los datos etiquetados. Al considerar la presencia de datos etiquetados en ambos dominios, la categorización nos sitúa en un área muy próxima al **MtL** pero una de las diferencias notables es que aquí existe una única tarea objetivo.

El área clave que aborda este trabajo se relaciona con que el ML tradicional, para asegurar el desempeño de los modelos, supone que los conjuntos de datos de entrenamiento y prueba comparten la **distribución de probabilidad**; suposición que muchas veces se incumple debido a distintos factores, como el arribo de una tarea perteneciente a un dominio novedoso o al alto grado de obsolescencia de los datos en ciertos dominios. Además, el proceso de etiquetado de nuevos datos es costoso y por consiguiente suele limitarse a un pequeño subconjunto, así como el desperdicio de grandes cantidades de datos previamente etiquetados en un dominio similar. En resumen, el objetivo de este trabajo es el desarrollo de metodologías que permitan clasificar con precisión datos de prueba novedosos, maximizando el uso de datos etiquetados en etapas de entrenamiento.

Teniendo en cuenta el contexto mencionado, este trabajo brinda un valioso aporte presentando un entorno denominado **TrAdaBoost**, que extiende el algoritmo de aprendizaje **AdaBoost** y además describe una serie de experimentos que demuestran que el método propuesto permite la construcción y aprendizaje de un modelo de clasificación de alta precisión, gracias a la transferencia de conocimiento utilizando una pequeña cantidad de datos novedosos etiquetados, conjuntamente con una gran cantidad de datos previamente etiquetados, evitando así el costoso proceso de etiquetado masivo, así como el potencial desperdicio de datos etiquetados disponibles, por los cuales ya se ha pagado el costo asociado al proceso.

Sea T_{D_D} el conjunto abundante de datos etiquetados de D_s cuya distribución difiere de la observada por los datos de D_T , la estrategia es maximizar su reuso, pero debido a que se desconocen las secciones útiles para la tarea de clasificación, se etiqueta una pequeña cantidad de datos en D_T ($T_{D_{SD}}$) con idéntica distribución objetivo con el objetivo de utilizarlos como referencia para hallar las secciones útiles de T_D , generando un clasificador que minimice el error predictivo con datos de prueba no etiquetados.

Debemos considerar que aunque ciertas secciones de los datos de entrenamiento pueden quedar desactualizadas, eso no evita la presencia de otras secciones utilizables para extraer y usufructuar el conocimiento valioso durante el entrenamiento de un clasificador de alta precisión. Para discernir esas secciones, se emplea una pequeña cantidad de nuevos datos etiquetados denominados **datos de entrenamiento de igual distribución** (TD_{SD}) que ayuden a determinar la utilidad de cada instancia de datos previos disponibles, posiblemente bajo diferente distribución, entonces los denominamos **datos de entrenamiento de distinta distribución** (TD_{DD}). El objetivo de **TrAdaBoost** es aprender un modelo de clasificación de alta **precisión** mediante la utilización de datos provenientes de ambos conjuntos de datos; iniciando su aprendizaje sobre TD_{SD} y luego seleccionando los datos de entrenamiento más útiles de TD_{DD} evitando incurrir en una pérdida promedio elevada, producto de la utilización indiscriminada de datos de entrenamiento adicionales, eventualmente inadecuados para la tarea objetivo.

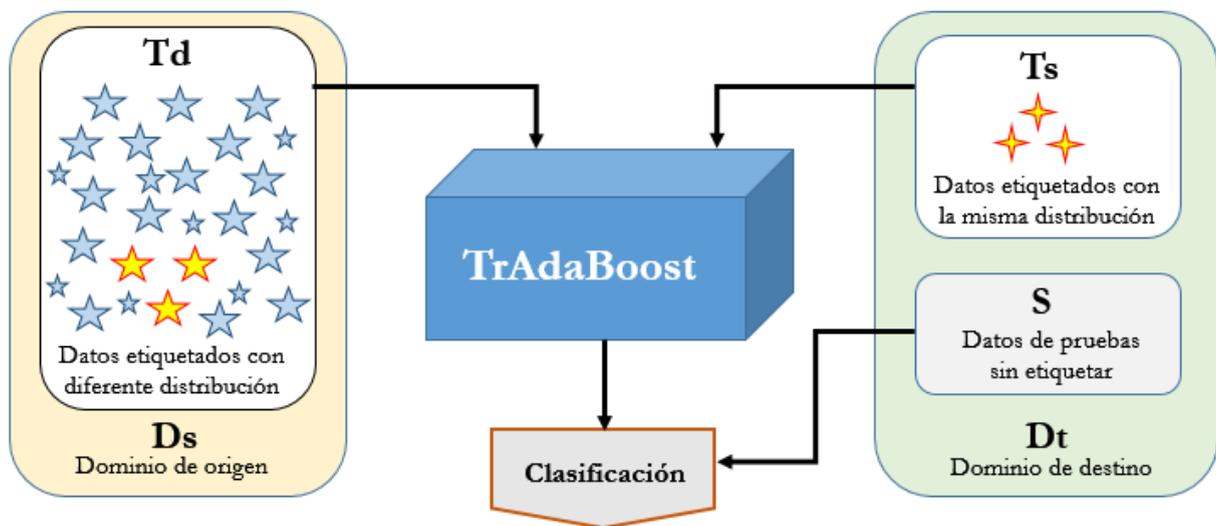


Figura 3.2: Esquema funcional del marco de trabajo TrAdaBoost
Figura de elaboración propia

El algoritmo **AdaBoost**, análogamente a los métodos tradicionales de ML, asume distribuciones idénticas de datos en fases de entrenamiento y prueba. En la extensión **TrAdaBoost**, aún se aplica a los datos de entrenamiento de misma distribución para construir la base del modelo pero para las instancias de entrenamiento de diferente distribución se añade un mecanismo para decrementar sus pesos a fin de debilitar su impacto. A medida que avanzan las iteraciones de entrenamiento, las **instancias** de TD_{DD} que se ajustan mejor a las de TD_{SD} tendrán pesos mayores mientras que las de diferente distribución tendrán pesos bajos. En síntesis, la idea principal es utilizar el impulso para filtrar los datos de entrenamiento de diferente distribución, ajustando automáticamente los pesos de instancias de entrenamiento; los datos restantes en TD_{DD} se tratan como adicionales y aumentan la confianza del modelo generado incluso cuando los datos de TD_{SD} escasean.

El análisis teórico desarrollado en este artículo demuestra que **TrAdaBoost** obedece primeramente a los datos de entrenamiento de la misma distribución de los datos objetivo (TD_{SD}) y luego determina las instancias de entrenamiento más útiles de diferente distribución (TD_{DD}) como datos de entrenamiento adicionales; permitiendo, la transferencia del conocimiento desde una distribución a otra mediante la aceleración de aprendices básicos y pudiendo concluir que la hipótesis final alcanzada por el marco de trabajo reduce progresivamente el error de los datos TD_{SD} y el error ponderado promedio en TD_{DD} que converge a cero gradualmente. El análisis experimental permite concluir que **TrAdaBoost** transfiere el aprendizaje de manera simple, alcanzando un elevado rendimiento y convergencia satisfactoria hacia el modelo deseado.

El análisis de este trabajo nos permite apreciar de qué manera la transferencia de conocimiento entre dominios con distribuciones de datos disímiles puede realizarse exitosamente, favoreciendo la incorporación de los datos más provechosos, reduciendo la de los menos útiles y filtrando aquellos dañinos para el modelo objetivo; evidenciando cómo el TL permite relajar ciertas restricciones poco realistas del ML tradicional.

3.4 - Aprendizaje autodidacta: TL desde datos sin etiquetar

R. Raina, A. Battle, H. Lee, B. Packer, A. Y. Ng [11]

Este trabajo se enmarca en la técnica de **TL inductivo (ITL)** ya que existe la presencia de datos etiquetados en D_T pero no así en D_S (ver **tabla 2.2** y **figura 2.4**). Además, dado que se intenta hallar una “adecuada” representación de características para el dominio objetivo que reduzca la diferencia entre dominios y el error de los modelos generados, corresponde al **enfoque representación de características**.

El desarrollo de este trabajo se centra en torno al interrogante: ¿es posible utilizar datos sin etiquetar de cualquier objeto y sencilla obtención, para mejorar el rendimiento de una tarea de clasificación supervisada?. Este enfoque se motiva en la observación de que muchos datos (imágenes, audios, textos, etc.) obtenidos aleatoriamente contendrán patrones subyacentes básicos similares a los de la tarea objetivo; por tanto, si logramos reconocer tales patrones extrayendo una representación sucinta de características de alto nivel de las entradas, podremos utilizar ese conocimiento para aprender la tarea objetivo, simplificando así su resolución.

El abordaje de este artículo permite presentar el novedoso marco de trabajo de ML denominado **aprendizaje autodidacta (StL)**, donde es posible utilizar datos sin etiquetar sobre tareas de clasificación supervisadas. Si bien este enfoque guarda similitudes con el **aprendizaje semi-supervisado (SsL)**, como la utilización tanto de datos etiquetados como sin etiquetar, presenta diferencias conceptuales, como el hecho que SsL supone que los datos sin etiquetar pueden etiquetarse con las mismas etiquetas que la tarea de clasificación principal y que presentan la misma distribución generativa. Además, el TL típicamente requiere más datos etiquetados (de costosa adquisición) de una tarea relacionada en lugar de datos sin etiquetar y generalmente transfiere conocimiento entre tareas supervisadas. Debido a que StL impone restricciones más laxas sobre la utilización masiva de datos sin etiquetar asequibles a relativo bajo costo, en la mayoría de las aplicaciones prácticas (como clasificación de imágenes, audio o texto) resultan ampliamente aplicables sobre multitud de problemas de aprendizaje y de manera más sencilla que los métodos típicos de **SsL** o **TL**. En síntesis, el estudio de **StL** intenta formalizar un marco de trabajo de **ML** que simplifique y abarate el proceso de aprendizaje sensiblemente (ver **figura 3.3**, donde las imágenes encuadradas representan datos etiquetados).

El enfoque seguido por **StL** consiste básicamente de dos etapas: el aprendizaje de una representación de alto nivel de abstracción desde los datos sin etiquetar; seguida de la aplicación de dicha representación sobre los datos etiquetados para su utilización en la tarea de clasificación objetivo. Un algoritmo de **StL** genera una hipótesis que trata de imitar fielmente la relación entrada-etiqueta representada por los datos de entrenamiento etiquetados, para luego probarla bajo la distribución generativa de los datos etiquetados. Concretamente, el algoritmo propuesto, comienza utilizando datos sin etiquetar descubriendo los elementos básicos compositivos aprendiendo representaciones de mayor grado de abstracción y luego aplica esa representación sobre los datos etiquetados del dominio destino, obteniendo una representación de más alto nivel de los mismos que le permiten optimizar el proceso de aprendizaje de la tarea objetivo.

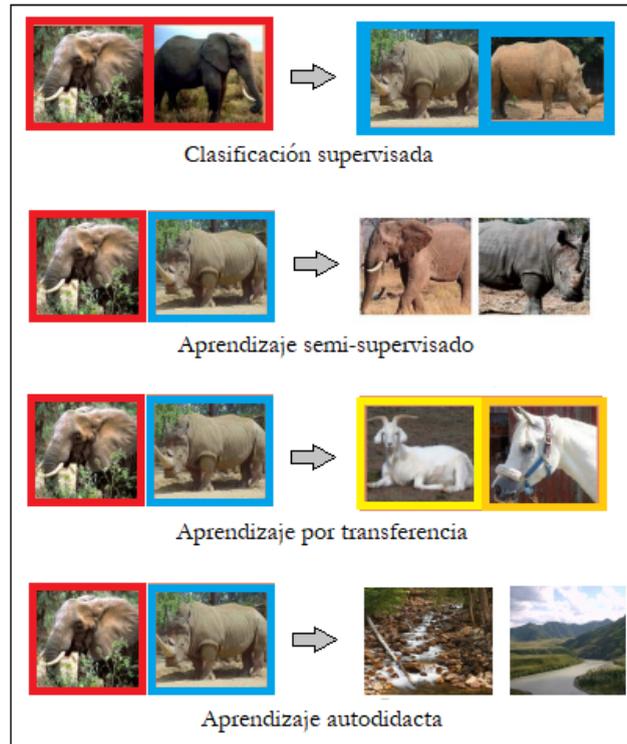


Figura 3.3: Distinción entre paradigmas según datos etiquetados

Figura adaptada de **Self-taught Learning: TL from Unlabeled Data** - Raina, Battle, Lee, Packer y Ng

Para el aprendizaje de representaciones bases de alto nivel a partir de los datos sin etiquetar se utiliza una versión modificada del **algoritmo de codificación dispersa (SCA)** aplicable a diferentes tipos de entradas, por ejemplo su aplicación sobre imágenes aprenderá a detectar bordes, contrastes y orientaciones (ver **figura 3.4**); sobre secuencias de audio, aprenderá a detectar patrones de frecuencias. Este trabajo además plantea comparativas entre diferentes algoritmos de generación de bases y sus combinaciones (**SCAs** y **PCAs**). De esa forma, ante cada **instancia** del conjunto de datos de entrenamiento etiquetado se capturan aspectos de alto nivel estructural, detectando patrones útiles coincidentes y generándose una serie de activaciones que capturan las características de alto nivel que posteriormente se utilizan como un nuevo conjunto de datos de entrenamiento para algoritmos tradicionales de clasificación supervisada (como **SVMs** o **GDA**s).

Analizando las dos últimas décadas de progreso del **ML**, se cree que el **StL** representa la extrapolación natural de una secuencia de formalismos propuestos por varios autores, desde el aprendizaje puramente supervisado, pasando por el **SsL** y arribando al **TL**, considerando la reducción progresiva de utilización de datos etiquetados, que acarrea elevados costos generativos, para dar lugar a la utilización de datos cada vez menos relacionados. Desde esta perspectiva, el **StL** también puede referirse como “transferencia no supervisada” o “TL desde datos sin etiquetar” (nombre que utiliza el trabajo analizado), ya que posibilita la mejora significativa de **modelos** sobre una gran variedad de dominios.

El tratamiento de este trabajo nos posibilita tener noción de la importancia que esta técnica del TL representa y el amplio espectro de dominios de aplicación que pueden sacar provecho de ella, considerando la generalidad de los datos que es posible utilizar para la extracción del conocimiento útil para potenciar el aprendizaje de tareas innovadoras a resolver.

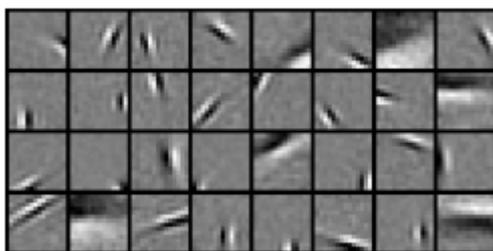


Figura 3.4: Bases aprendidas por SCA para una imagen

Figura adaptada de **Self-taught Learning: TL from Unlabeled Data** - Raina, Battle, Lee, Packer y Ng

3.5 - TL mediante mapeo estructural local de modelos múltiples

Jing Gao, Wei Fan, Jing Jiang y Jiawei Han [15]

La transferencia de conocimiento a través de dominios que presentan diferentes distribuciones es un problema importante y parcialmente investigado en el campo de la **minería de datos**. Siendo parte de la técnica de **transferencia inductiva (ITL)** y más precisamente del enfoque de **transferencia de parámetros** ya que se asume que los dominios involucrados comparten parámetros o distribuciones previas de hiper parámetros plausibles de ser utilizados para transferir conocimiento útil entre ellos. Este trabajo aprovecha los poderes predictivos de modelos entrenados sobre diferentes dominios con algunos datos etiquetados y diversos algoritmos de aprendizaje, proponiendo un **marco de trabajo conjunto ponderado localmente (LWE)** para la transferencia de conocimiento hacia un dominio diferente (aunque relacionado) sin ejemplos etiquetados.

Desafortunadamente, las técnicas tradicionales de transferencia supervisada (**STL**) no hacen un buen trabajo en estas situaciones ya que su eficacia depende del grado de similitud entre las distribuciones propias de los conjuntos de datos de entrenamiento y prueba, y el tratamiento de muestras **iid** procedentes de la misma distribución. Además, diferentes modelos presentan distintos conocimientos, producto del sesgo inductivo de cada método de aprendizaje y las distribuciones entre dominios de entrenamiento, presentando diferentes grados de efectividad sobre distintos dominios y regiones del dominio de prueba. Por lo tanto, se propone **LWE** para la integración del conocimiento proveniente de múltiples modelos base (clasificadores supervisados) generados a partir de distintos dominios de entrenamiento y su posterior aplicación sobre un modelo de clasificación unificado sobre el dominio de prueba, en pos de mayor efectividad de aprendizaje.

Cuando los modelos base conllevan conceptos contradictorios o inciertos ante un ejemplo de prueba, cobra importancia la selección del más adecuado, es decir, que mejor represente la distribución objetivo real subyacente para dicha instancia, con la esperanza de que la precisión del enfoque combinado supere la de cualquier modelo individual que actúe aisladamente. LWE propone un modelo gráfico de aproximación y comparativa de vecindad de ponderaciones óptimas basado en estructuras conjuntas de dominios y una etapa de ajuste de predicción para propagar etiquetas de ejemplos cercanos ante modelos base inconsistentes. Desarrollando una metodología que beneficia los clasificadores cuya estructura local mapeada se asemeja más a la del ejemplo de prueba actual. Para ello utiliza una matriz de ponderaciones para cada ejemplo de prueba y a partir de su localidad con otros ejemplos de entrenamiento, utilizando los vectores de pesos más adecuados y combinando las predicciones de diferentes modelos óptimos para ese caso puntual, minimizando el error y en consecuencia, maximizando la precisión de clasificación.

De este modo, ya que diferentes ejemplos son predichos con mayor exactitud por diferentes modelos de entrenamiento y no existe un modelo que prediga correctamente a todos, resulta vital un enfoque de ponderación por ejemplos en lugar de global en términos de precisión de clasificación. Para la selección del modelo óptimo según el ejemplo actual, se utiliza una metodología comparativa entre cada grafo de modelo base y un grafo objetivo con la estructura de agrupamiento de todos los ejemplos de prueba y se utiliza el porcentaje de coincidencias entre vecinos para aproximar la precisión de cada modelo base alrededor del ejemplo actual. Este enfoque comparativo de consistencia estructural entre modelos base y objetivo, posibilita la selección del modelo más apropiado ante cada ejemplo de prueba. Ante modelos base en conflicto y donde ninguno predice satisfactoriamente el ejemplo actual, se abandona la información de los clasificadores supervisados para construir un clasificador no supervisado basado en la estructura de vecindad; permitiendo así, la predicción acertada de ejemplos en áreas inciertas y su correcto etiquetado.

El marco de trabajo propuesto se evaluó sobre cuatro tareas, una ficticia (campo de juguetes) y tres realistas (filtrado de correo electrónico no deseado, categorización de textos y detección de intrusiones en redes); comparándolo con cuatro métodos de modelado individual de referencia (**WNNa**, **LR**, **SVM** y **TSVM**) dada su estrecha relación con el problema. Además, se incluyen tres métodos comparativos: **SMA**, donde las predicciones son combinadas utilizando pesos uniformes, **pLWE**, donde no existe la etapa de ajuste y **LWE**, que involucra combinación de clasificadores y ajuste basado en estructura local.

Los resultados experimentales demuestran que el enfoque propuesto supera los algoritmos de modelo individual supervisados, semi-supervisados y de conjunto promediado simple. Alcanzando rangos de mejora de hasta 30 %. Debido a la discrepancia entre distribuciones de entrenamiento y prueba, los métodos de modelo individual estudiados usualmente presentan un pobre rendimiento (con **precisión** cercana al 70 % y **MSE** superior a 0.1), ya que sólo se sustentan en la información etiquetada del dominio origen sin esforzarse su selección útil ni la transferencia del conocimiento adecuado hacia el dominio destino.

En el esquema de ponderación, la etapa de ajuste filtra la información “dañina”, incrementando la fiabilidad de las predicciones, especialmente sobre conjuntos de pruebas donde los dominios de entrenamiento parecen estar en conflicto entre sí y por lo tanto entrenados a partir de la unión de ambos dominios y el promediado simple de ambos modelos resulta en una precisión cercana al 60 %, mientras que **LWE** la eleva al 96.36 % gracias a la selección de información útil desde ambos modelos. Además, se observa que la precisión del peor modelo individual es cercana al 56 %, pudiendo **SMA** incluso degradarla al 54 %, ya que la combinación ingenua de modelos no concordantes (con grandes áreas de incertidumbre) producen la propagación de información incorrecta reduciendo la precisión de clasificación del modelo resultante.

En síntesis, la precisión final de estos métodos de promediado simple así como los semi-supervisados dependerán en gran medida de los datos y los modelos de entrenamiento intervinientes en el problema. Incluso, podemos notar que incluso basándonos en esos clasificadores tan débiles, **LWE** aún puede mejorar la precisión al 80 %. Para más detalles, observar la **tabla 3.1** donde podemos apreciar que la dificultad de **TL** está marcada por la discrepancia desconocida entre las distribuciones de entrenamiento y prueba de los conjuntos de datos, evidenciado en los bajos índices de precisión asociados conforme mayor es el grado de discrepancia en las distribuciones.

Precisión												
Métodos	Filtrado de correo no deseado			20 Grupo de Noticias						Reuters		
	U00	U01	U02	C vs S	R vs T	R vs S	S vs T	C vs R	C vs T	O vs Pc	O vs PI	Pc vs PI
WNN	0.7680	0.7888	0.8696	0.6554	0.5938	0.7942	0.7557	0.8926	0.9341	0.7058	0.6520	0.5685
LR	0.7060	0.7528	0.8500	0.7349	0.7217	0.7885	0.7904	0.8334	0.9176	0.7355	0.7122	0.5565
SVM	0.6604	0.7288	0.7844	0.7118	0.6824	0.7816	0.7577	0.8156	0.9389	0.6934	0.6998	0.5694
SMA	0.7416	0.8012	0.8768	0.7272	0.6845	0.7980	0.7806	0.8563	0.9348	0.7339	0.7008	0.5685
TSVM	0.8352	0.8512	0.9528	0.7697	0.8995	0.8996	0.8559	0.8964	0.8826	0.7380	0.6989	0.5843
pLWE	0.8584	0.8820	0.9520	0.7872	0.7217	0.8845	0.8330	0.9193	0.9664	0.7694	0.7008	0.5972
LWE	0.8908	0.8844	0.9820	0.9744	0.9923	0.9823	0.9692	0.9816	0.9890	0.7967	0.7304	0.6852

Error cuadrático medio (MSE)												
Métodos	Filtrado de correo no deseado			20 Grupo de Noticias						Reuters		
	U00	U01	U02	C vs S	R vs T	R vs S	S vs T	C vs R	C vs T	O vs Pc	O vs PI	Pc vs PI
WNN	0.1836	0.1713	0.1003	0.2775	0.2968	0.1575	0.1978	0.0851	0.0525	0.2462	0.3055	0.3774
LR	0.1944	0.1672	0.1013	0.2057	0.2036	0.1567	0.1624	0.1340	0.0613	0.2190	0.2444	0.3900
SVM	0.2374	0.1890	0.1489	0.2140	0.2353	0.1644	0.1826	0.1360	0.0453	0.2217	0.2230	0.2827
SMA	0.1556	0.1337	0.0870	0.2030	0.2183	0.1349	0.1614	0.0979	0.0430	0.1987	0.2318	0.3049
TSVM	0.1428	0.1394	0.0814	0.1749	0.1080	0.1128	0.1281	0.1198	0.1061	0.2250	0.2128	0.2688
pLWE	0.1218	0.1012	0.0550	0.1795	0.2027	0.1029	0.1399	0.0699	0.0302	0.1845	0.2333	0.3000
LWE	0.0988	0.1022	0.0333	0.0965	0.1409	0.0384	0.0534	0.0308	0.0140	0.1678	0.2120	0.2091

Tabla 3.1: Comparación de rendimiento de métodos en diferentes conjuntos de datos

Tabla adaptada de **Knowledge Transfer Via Multiple Model Local Structure Mapping - Gao, Fan, Jiang y Han**

Los resultados experimentales presentes en este trabajo demuestran que el marco de trabajo propuesto **LWE** identifica exitosamente el conocimiento útil y relevante de cada modelo base de entrenamiento para predecir correctamente los ejemplos de pruebas, realizando un uso apropiado del TL que incrementó considerablemente la precisión del modelo resultante. Además se efectúan ponderaciones conjuntas óptimas, promoviendo la flexibilidad de modelos base, al no requerir que sean diseñados específicamente para TL.

3.6 - Mapeo y revisión de redes lógicas de Markov para TL

Lilyana Mihalkova, Tuyen Huynh y Raymond Mooney [13]

Este trabajo considera el **TL** mediante **redes lógicas de Markov (MLNs)**, presentando un sistema de transferencia que mapea autónomamente los predicados desde una MLN de origen hacia el dominio destino y luego revisa la estructura mapeada a fin de incrementar su precisión. Correspondiendo a la técnica de **transferencia inductiva (ITL)** y perteneciente al enfoque de **transferencia de conocimiento relacional** ya que se presupone la existencia de alguna relación entre los datos del dominio origen y destino involucrados en el proceso de transferencia, representando dicha relación el conocimiento a ser transferido.

El **algoritmo de transferencia mediante mapeo y revisión automática (TAMaR)** se motiva en el hecho de que si dos dominios se encuentran relacionados, existen mapeos o correspondencias que conectan sus entidades y relaciones; siendo posible efectuar la transferencia de conocimiento mediante MLNs. Siguiendo esa línea, TAMaR intenta utilizar una MLN aprendida sobre el dominio origen para optimizar el aprendizaje de otra MLN de dominio destino. Básicamente, se trata de un algoritmo en dos fases; donde inicialmente se construye un mapeo desde una MLN de origen hacia D_T y luego se revisa la estructura mapeada allí mediante un algoritmo de **programación lógica inductiva (ILP)** denominado **FORTE**.

El aprendizaje de una MLN considera dos aspectos fundamentales: la estructura, que requiere un procesamiento computacionalmente intensivo y los pesos, cuyo procesamiento es relativamente rápido y con gran potencial beneficioso para la transferencia. Por su parte, la transferencia estructural entre MLNs, consiste de dos subtarear:

1. **Mapeo de predicados**, donde mediante un conjunto de predicados se describen las entidades de datos de los dominios relacionales de origen y destino (parcial o totalmente distintos) y sus relaciones mediante la lógica de primer orden. Por lo tanto, la primera tarea de transferencia es establecer un mapeo o asignación de correspondencias desde los predicados del dominio origen hacia el de objetivo permitiendo la traducción de cláusulas sobre este último. El objetivo es hallar la mejor correspondencia entre redes de origen y destino; cuya calidad es medida según el rendimiento de la MLN mapeada sobre los datos destino.

Existen dos enfoques principales de mapeo: el **general**, usualmente prohibitivo ya que intenta establecer un mapeo para cada predicado de origen a un predicado destino y traducir la red de origen completa; y el **local**, que encuentra el mejor mapeo para cada cláusula origen generando un mapeo sólo para los predicados que aparecen en esa cláusula, lo cual es escalable ya que el número de predicados de una cláusula es reducido respecto a los de todo el dominio origen.

2. **Refinamiento teórico o revisión de la estructura mapeada**, las cláusulas traducidas en la etapa de mapeo podrían no ser completamente precisas y requerir revisiones a fin de lograr un mejor ajuste sobre los datos del dominio destino.

Consta de tres etapas: **autodiagnóstico**, donde se inspeccionan las cláusulas de la MLN origen determinando partes inexactas y evaluando cuáles deberían acortarse, alargarse o mantenerse intactas; **actualización**, donde se efectúan las acciones pertinentes sobre cada cláusula, con un espacio de búsqueda restringido; y el **descubrimiento de nuevas cláusulas**, utilizando **RPF** se descubren nuevas cláusulas en D_T evaluadas mediante **WPLL** y aquellas que mejoran la valuación general se incorporan a la MLN.

Este trabajo analiza el rendimiento de tres sistemas: **SrcKD**, un sistema de aprendizaje ejecutado desde cero sobre D_T ; **TrKD**, un sistema de transferencia usado para revisión de la estructura de origen traducida a D_T mediante el procedimiento de mapeo automático de predicados; y **TAMaR**, el sistema de transferencia propuesto mediante mapeo y revisión automática. Así mismo se utilizaron cuatro dominios relacionales del mundo real para pruebas: **IMDB**, una base de datos internacional de películas; **UW-CSE**, una base de datos de personas del departamento académico de ciencias informáticas; **WebKB**, entidades del conjunto de datos del departamento informático universitario; **UW-KB**, una base de conocimiento construida manualmente desde UW-CSE, usada como MLN de origen. Finalmente, se utilizaron dos métricas: **curva de recuperación de precisión (AUC)** para indicar que tan bien el algoritmo predice instancias positivas y la **probabilidad de registro condicional (CLL)** para determinar la calidad de las predicciones efectuadas. Por su parte, las curvas de aprendizaje se han resumido mediante dos estadísticas: **relación de transferencia (TR)** que indica la relación entre áreas bajo las curvas de aprendizaje indicando el grado de mejora de la transferencia respecto del aprendizaje desde cero y el **porcentaje de mejora de 1 mega ejemplo (PI)**, que mide el grado en que el TL mejora la precisión respecto al aprendizaje desde cero a partir de observar un único mega-ejemplo en D_T .

Como se aprecia en la **tabla 3.2**, en términos de AUC, apreciamos que ambos sistemas de transferencia muestran mejoras sobre un aprendiz ingenuo (desde cero) aunque ningún aprendiz de transferencia supera notablemente a otro. En términos de CLL, apreciamos que la transferencia siempre mejora el aprendizaje desde cero y el rendimiento de TAMaR supera notablemente al de TrKD.

Caso Experimental	AUC				CLL			
	TR		PI		TR		PI	
	Trkd	TAMAR	Trkd	TAMAR	Trkd	TAMAR	Trkd	TAMAR
WebKB → IMDB	1.51	1.55	50.54	53.90	1.41	1.46	51.97	67.19
UW-CSE → IMDB	1.42	1.66	32.78	52.87	1.33	1.56	49.55	69.28
UW-KB → IMDB	1.61	1.52	40.06	45.74	1.21	1.44	30.66	58.62
WebKB → UW-CSE	1.84	1.78	47.04	37.43	1.17	1.36	19.48	32.69
IMDB → UW-CSE	0.96	1.01	-1.70	-2.40	1.62	1.67	34.69	54.02

Tabla 3.2: Métricas TR y PI en términos de AUC y CLL sobre ScrKD
 Tabla adaptada de **Mapping and Revising Markov Logic Networks for Transfer Learning**

Si bien ambos sistemas de transferencia son más rápidos que ScrKD, TAMaR entrena más rápido y considera menos cláusulas candidatas que TrKD, que a su vez considera más candidatos que ScrKD pero requiere menos tiempo de entrenamiento ya que cada candidato es evaluado sobre menos datos. A partir de la **tabla 3.3**, donde se describen los tiempos (en minutos) de entrenamiento promedio y entre paréntesis el tiempo promedio necesario para construir el mapeo de predicados (en color verde expresando segundos), además de los porcentajes promedio de tiempo en color azul). Podemos apreciar que los tiempos empleados por algoritmos de mapeo representan una pequeña fracción del tiempo de entrenamiento.

A diferencia del enfoque **MtL**, donde el algoritmo es simultáneamente proporcionado a todos los dominios durante el entrenamiento y construye una estructura compartida por los modelos, lo novedoso del enfoque propuesto aquí es que el aprendizaje encuentra los dominios uno a la vez. Otro factor novedoso del trabajo es que el mapeo de predicados propuesto se basa en la compatibilidad de predicados entre dominios sin requerir reglas de mapeo y son construidos de forma completamente automática.

El sistema de TL propuesto (**TAMaR**) mapea autónomamente una **MLN** desde D_S hacia D_T y posteriormente efectúa una revisión estructural a fin de mejorar aún más su rendimiento. Los resultados empíricos presentados demuestran que TAMaR mapea con éxito el conocimiento transferido para utilizarlo en la mejora de precisión del modelo aprendido, reduciendo significativamente los tiempos y datos necesarios durante la fase de entrenamiento. Debido a que TAMaR diagnostica la estructura mapeada y sólo se limita a revisar sus secciones incorrectas, aprende significativamente más rápido y considera muchos menos candidatos que el mejor aprendiz **MLN** de la actualidad.

Experimento	ScrKD	TrKD	TAMAR
WebKB → IMDB	62.23'	32.20' (0.89" - 0.046%)	11.98' (0.89" - 0.12%)
UW-CSE → IMDB	62.23'	38.09' (9.18" - 0.39%)	15.21' (9.10" - 0.99%)
UW-KB → IMDB	62.23'	40.67' (9.98" - 0.41%)	6.57' (9.99" - 2.53%)
WebKB → UW-CSE	1127.48'	720.02' (6.71" - 0.016%)	13.70' (6.75" - 0.82%)
IMDB → UW-CSE	1127.48'	440.21' (42.48" - 0.16%)	34.57' (42.28" - 2.04%)

Tabla 3.3: Tiempos de entrenamiento y construcción de mapeos
 Tabla adaptada de **Mapping and Revising Markov Logic Networks for Transfer Learning**

El abordaje de este novedoso método nos permite observar de que forma los algoritmos intentan hacer un uso cada vez más restringido de los datos a fin de minimizar procesamiento y tiempos requeridos, así como establecer optimizaciones mediante el refinamiento de las secciones estructurales que no son de utilidad para el proceso de transferencia, lo que permite establecer progresivamente métodos superadores en el área.

3.7 - Aprendizaje y evaluación de clasificadores bajo el sesgo de selección de la muestra Bianca Zadrozny [16]

Este trabajo se encuentra bajo el **enfoque de transferencia de instancias** perteneciente a la **técnica de transferencia transductiva (TTL)** ya que trabaja modificando el **dominio** de origen correspondiente a los datos de entrenamiento, el cual admite la presencia de datos etiquetados, para su aplicación sobre un dominio objetivo que carece de datos etiquetados, como indica la **tabla 2.2**; y considerando además, idénticas las tareas de origen y destino involucradas, así como dominios diferentes como indica la **tabla 2.1**. Se considera el caso donde los espacios de características entre dominios coinciden pero no así las distribuciones de probabilidad asociadas (ver **sección 2.6.2**), escenario denominado transferencia homogénea (**HomTL**).

En muchas situaciones prácticas donde se desea hallar un clasificador que prediga ejemplos pertenecientes a un conjunto de datos con determinada **distribución** subyacente. Para lograrlo, lo ideal sería recolectar **muestras insesgadas** (aleatorias) de la población que representen fielmente dicha distribución, pero en ocasiones este tipo de recopilación se dificulta al no disponer el control del proceso de recolección de datos o incluso por disponer de muestras previamente recolectadas. Los métodos clasificatorios tradicionales suelen asumir (implícita o explícitamente) que el modelo aprendido realizará predicciones utilizando datos de entrenamiento y prueba con idénticas distribuciones y que además los ejemplos de evaluación son extraídos de forma aleatoria e independiente; pero cuando esas premisas se incumplen, emerge el problema del **sesgo de selección de la muestra (SSB)**, que el modelo generado debe solventar para lograr un alto rendimiento.

Este trabajo representa un valioso aporte al campo del TL proporcionando una categorización de clasificadores novedosa (locales y globales) en lugar de la tradicional (discriminativos y generativos). Los clasificadores locales no son sensibles al SSB, dependiendo su salida exclusivamente de las distribuciones locales de las entradas respecto a cada etiqueta ($P(y|x)$), por el contrario, los globales sí son afectados ya que su salida depende también de la distribución global del espacio de características ($P(x)$). Los resultados analíticos de cada tipo de clasificador analizado en este trabajo se resumen en la **tabla 3.4**.

Método de clasificación		Categorización	Afectado por el SSM
Bayesiano	Estricto	Local	NO
	Ingenuo	Global	SÍ
Regresión logística		Local	NO
Árbol de decisión		Global	SÍ
SVM	Margen rígido	Local	NO
	Margen flexible	Global	SÍ

Tabla 3.4: Métodos de clasificación y su afectación ante el SSM
Tabla de elaboración propia

Sea el dominio de la muestra $D(X, Y, S)$, siendo los espacios: X de características, Y de etiquetas y S de selección binaria de muestras) decimos que si “ s ” es independiente de “ y ” dado “ x ”, ($P(s|x, y) = P(s|x)$), la muestra seleccionada estará sesgada, y este sesgo sólo depende del vector de características x . Si bien pueden existir otros tipos de sesgos, el presente trabajo se focaliza en este caso dado que es el más realista.

Este trabajo realizó experimentos con algoritmos de **regresión logística**, **Bayesiano** (ingenuo), **árbol de decisión** (C4.5) y **SMSVM**; detallando los resultados en la **figura 3.5** donde se aprecia la evolución de la tasa de error conforme aumenta el conjunto de entrenamiento, correspondiendo líneas superiores punteadas en color rojo a datos sesgados, continuas inferiores en color verde a insesgados y las barras al error estándar.

En base a los resultados experimentales apreciamos que los clasificadores de regresión logística, no se ven afectados por el SSB. En cuanto a los métodos generativos (como Bayesianos estrictos) sabemos que su comportamiento es generalmente independiente de $P(x)$, aunque esto no se cumple para el caso estudiado (ingenuo) muy sensibles al SSB evidenciado en el crecimiento de la tasa de error. Con respecto a los árboles de decisión, contrariamente a lo esperado, observamos que se desempeñan bien bajo el SSB, lo cual podría deberse al hecho de que aunque la elección de las particiones empleadas es sesgada, no así las estimaciones de clases en las hojas. Finalmente en el caso de **SVMs** de margen suave, vemos que la tasa de error utilizando un conjunto de entrenamiento sesgado decrece, sin embargo, la diferencia entre las tasas de error con muestras sesgadas e insesgadas no lo hace, siendo indicador de que los **SMSVMs** se ven afectados por el SSB.

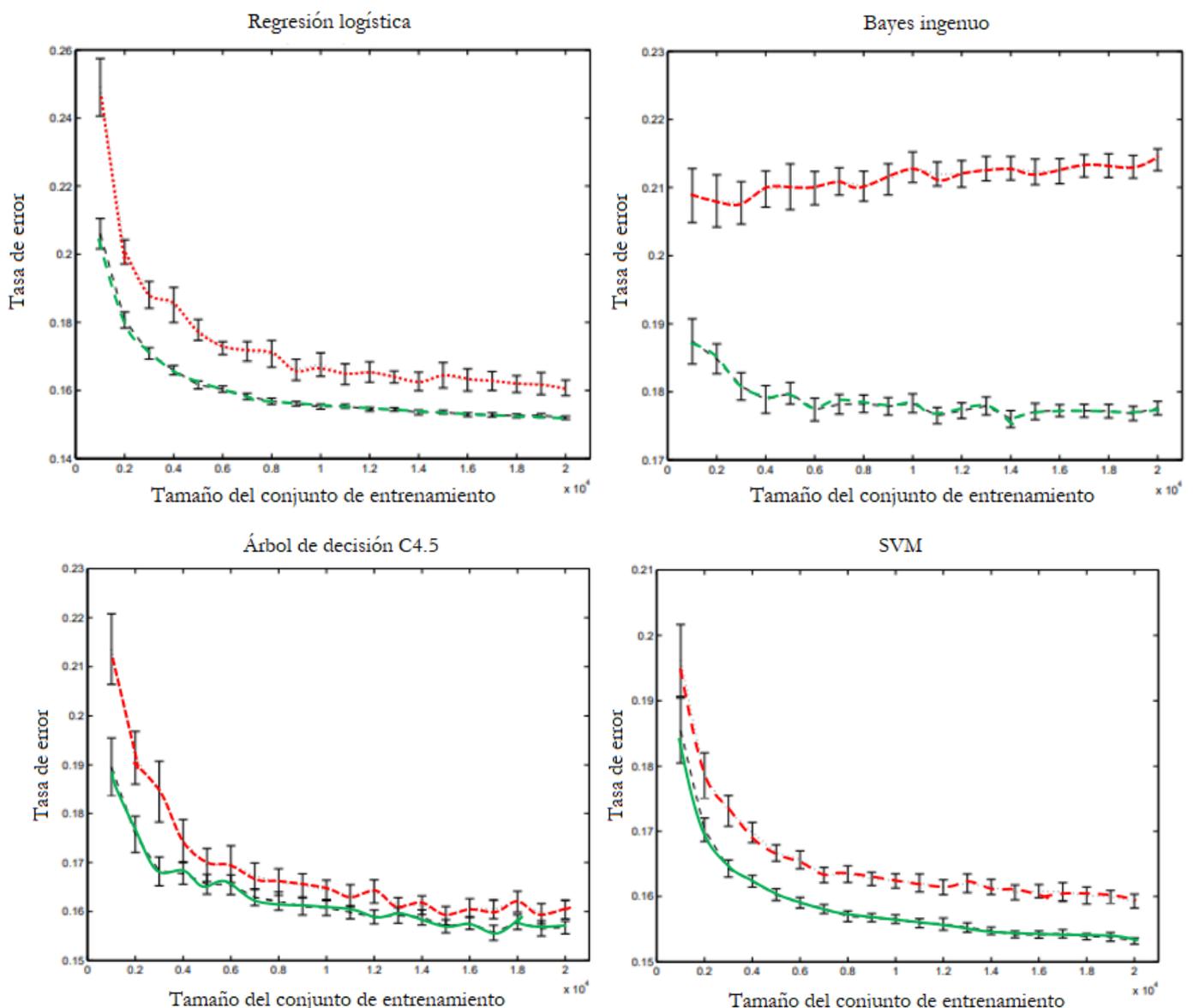


Figura 3.5: Tasas de error con datos de entrenamiento sesgados e insesgados
 Figura adaptada de **Learning and Evaluating Classifiers under Sample Selection Bias - Zadrozny**

Un aspecto novedoso, es la categorización de los clasificadores evaluados, en locales (insensibles al SSB) y globales (sensibles al SSB) que pueden hacer uso de datos sin etiquetar, cobrando relevancia el método de corrección propuesto. Otro aporte sobresaliente de este trabajo es la presentación de un método de corrección del **SSB** aplicable a cualquier clasificador, que intenta minimizar el valor esperado de la función de pérdida de la distribución de ejemplos y luego corregirla mediante la reponderación de cada muestra seleccionada con el fin de que su distribución se centre cercana al valor previamente estimado de la muestra completa, seguido de la aplicación del clasificador sobre la muestra corregida. De esta forma se logra que los clasificadores categorizados como globales y sensibles al SSB (como Bayes ingenuo, SMSVMs y árboles de decisión) se comporten tan bien como los locales, insensibles al SSB (como regresión logística o SVMs de margen rígido).

La importancia principal de este trabajo sobre el campo del **TL** consiste en el estudio analítico y experimental del impacto del **SSB** sobre el aprendizaje y evaluación de varios métodos de clasificación tradicionales, además de la presentación de un método correctivo muy útil en fase de evaluación de modelos.

3.8 - Aprendizaje por transferencia mediante reducción de dimensionalidad

Sinno Jialin Pan, James T. Kwok y Quiang Yang [40]

Este trabajo se ubica dentro de la técnica de **transferencia transductiva (TTL)** donde las tareas involucradas en el proceso de transferencia coinciden pero los dominios involucrados difieren (ver **tabla 2.1**). Más específicamente corresponde al enfoque de transferencia basado en **representación de características transductiva**, ámbito en el que la mayoría de los métodos existentes asumen que las distribuciones de probabilidad marginal de los datos de origen y destino pueden diferir, mientras que la distribución de probabilidad condicional respecto de las etiquetas, son similares. Además, en el contexto del TL, uno de los problemas recurrentes se da al intentar aprender un modelo cuando los datos de entrenamiento y validación (pertenecientes a D_S) y los de pruebas (pertenecientes a D_T) poseen diferentes **distribuciones o espacios de características**, lo que clasifica al método actual como perteneciente al **HetTL**. Uno de los enfoques resolutivos (el aplicado aquí) intenta aprender una estructura de características común que sirva de puente entre dominios, posibilitando el proceso de transferencia de conocimiento entre ellos.

El objetivo de este trabajo y su principal contribución al área del TL se centra en la presentación de un novedoso método que intenta descubrir un espacio latente de características compartido entre dominios que no se vea afectado por factores dinámicos y donde las distribuciones marginales correspondientes a los datos en los dominios implicados se encuentren estrechamente relacionadas, lo que permitirá disponer de un espacio latente de baja **dimensionalidad** que minimice la distancia entre esas distribuciones, sirviendo como enlace para la transferencia y por consiguiente, otorgando mayor eficiencia al proceso. La aplicación de esta metodología es posible incluso cuando D_T carece de datos etiquetados, permitiendo aprender un clasificador o regresor de alto rendimiento que saque el mayor provecho de los datos de entrenamiento etiquetados de D_S .

El enfoque reducción de dimensionalidad propuesto, presenta dos etapas:

1. La definición del novedoso método de **incrustación máxima de discrepancia media (MMDE)**, que aprende un espacio latente de baja dimensionalidad común a ambos dominios.
2. La aplicación de un algoritmo tradicional de **ML** (clasificador o regresor) entrenado desde la proyección de entradas aprendidas sobre el espacio previo para efectuar predicciones sobre D_T .

Se experimentó sobre dos dominios de prueba: localización mediante señales **WiFi** en entornos cerrados y clasificación de textos. Sobre el primero, se utilizaron diferentes modelos de regresión (**RLSRs**, **SVRs**, y **LapRLSRs**) con el objetivo de comprobar que esos modelos tradicionales con la ayuda de MMDE efectivizan la resolución de problemas de TL; evidenciándose que los correspondientes modelos entrenados sobre el espacio latente de dimensionalidad reducida obtienen un mayor rendimiento que aquellos entrenados sobre el **espacio de características** original, esta mejora no sólo se debe a la eliminación del ruido en los datos originales, sino además a la equiparación de distribuciones por parte de MMDE (ver **figura 3.6**).

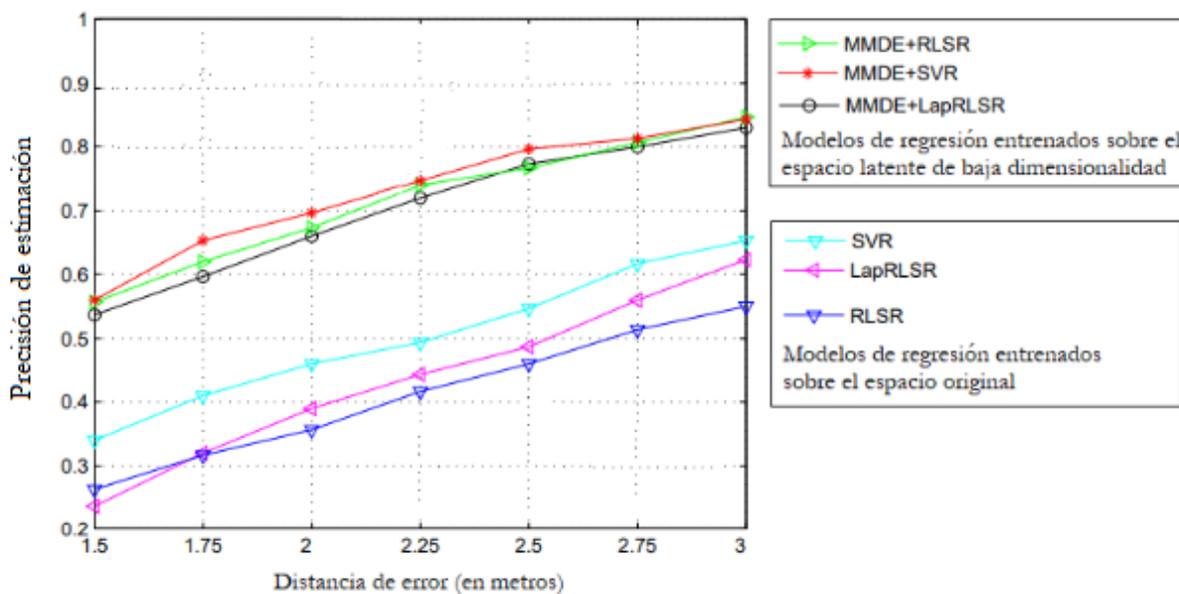


Figura 3.6: Comparación de precisión en métodos asistidos por MMDE
 Figura adaptada de **Transfer Learning via Dimensionality Reduction - Pan, Kwok y Yang**

Sobre el dominio de pruebas de clasificación de textos se utilizaron diferentes modelos de clasificación (**SVMs**, **TSVMs**) para verificar la capacidad de transferencia del algoritmo propuesto, pudiéndose apreciar que los modelos entrenados en el espacio latente aprendido por MMDE logran una mayor precisión que los entrenados sobre el espacio original y confirmando así, el éxito de MMDE en entornos de TL (ver **tabla 3.5**).

Conjunto de datos	Documentos		Palabras	SVM		TSVM	
	Dom. origen	Dom. destino		Original	MMDE	Original	MMDE
Personas vs. lugares	1079	1088	8000	0.52 (0.039)	0.65 (0.021)	0.55 (0.025)	0.67 (0.036)
Organizaciones vs. personas	1239	1210	9729	0.66 (0.021)	0.72 (0.034)	0.70 (0.026)	0.73 (0.033)
Organizaciones vs. lugares	1016	1046	8568	0.67 (0.025)	0.71 (0.021)	0.71 (0.035)	0.74 (0.036)

Tabla 3.5: Precisión de clasificadores entrenados sobre diferentes espacios de características
 Tabla adaptada de **Transfer Learning via Dimensionality Reduction - Pan, Kwok y Yang**

El aporte significativo que este trabajo realiza al campo del TL gira en torno al desarrollo de una novedosa técnica basada en el aprendizaje de un espacio latente de baja dimensionalidad mediante la proposición del algoritmo **MMDE** que permite la transferencia de conocimiento entre diferentes dominios. La experimentación efectuada demostró que la solución desarrollada incrementa significativamente el rendimiento sobre algoritmos tradicionales de ML.

La relevancia del tratamiento de este artículo en la tesina, se sustenta en el amplio espectro de situaciones realistas donde esta metodología puede aplicarse y su utilidad en dominios donde no se dispone de datos previamente etiquetados o asimismo cuando la dimensionalidad elevada sea problemática.

3.9 - Reducción de dimensionalidad transferida

Zheng Wang, Yangqiu Song y Changshui Zhang [20]

Este trabajo se enmarca dentro de la técnica de **TL no supervisada (UTL)** donde tanto las tareas como los dominios involucrados en el proceso de transferencia difieren, aunque se encuentran relacionados (ver **tabla 2.1**); focalizando en la resolución de tareas de **aprendizaje no supervisado** sobre un **dominio** objetivo mediante técnicas como **agrupamiento, reducción de dimensionalidad y estimación de densidad** (ver **figura 2.4**) y conforme a situaciones donde no se dispone de datos etiquetados en ningún dominio durante el período de entrenamiento (ver **tabla 2.2**). Este artículo se ubica más específicamente dentro del **enfoque de transferencia de representación de características** (ver **tabla 2.3**), ya que intenta hallar la representación más adecuada que contribuya a reducir la diferencia entre dominios y el error de los modelos generados, en este caso particular, mediante la aplicación de la técnica de **agrupamiento**.

El aporte relevante de este trabajo gira en torno a la presentación del novedoso problema de **reducción de dimensionalidad transferida (TDR)** que consiste en una reducción de dimensionalidad discriminativa y no supervisada, asistida mediante el conocimiento previo relacionado a otras clases; acompañado por la proposición del **algoritmo de análisis discriminativo transferido (TDA)** que aborda **TDR** mediante la técnica de agrupamiento para generar etiquetas de pseudo-clases a partir de los datos sin etiquetar del dominio destino y extraer información discriminativa desde éstos mediante la técnica de reducción de dimensionalidad aplicada conjuntamente con los datos etiquetados generados para la obtención de un subespacio representativo. Entonces, gracias a la generación de datos etiquetados y el uso de datos sin etiquetar se encuentra simultáneamente un agrupamiento óptimo y un subespacio discriminativo adecuado.

Muchas aplicaciones de **ML** demandan tratar con elevada dimensionalidad pero se debe comprender que los datos residen en un espacio de dimensionalidad reducida y hallar dicha representación compacta y simplificada posibilitará alcanzar una mayor eficiencia, sin pérdida de expresividad, aceleración y robustez en fases de modelado subsecuentes. Motivando el desarrollo de enfoques de **reducción de dimensionalidad** bajo los paradigmas supervisados (como **LDA**), semi-supervisado y no supervisado (como **LPP** y **PCA**).

Para obtener esa estructura discriminativa, los métodos supervisados intentan encontrar una transformación que minimice la dispersión en cada clase y al mismo tiempo maximice la dispersión entre clases. El método más reciente utiliza el algoritmo **DisK-medias**, uno de los más simples y populares para resolver el problema de agrupación a partir de un conjunto de dato sin etiquetar y aprovechar su estructura discriminativa; superando al framework **PCA+K-medias** y otros similares, pero su eficiencia se restringe a ciertas distribuciones y puede eventualmente sufrir estancamientos; para solventarlo, se considera la incorporación de información proveniente de otros dominios relacionados (mismo dominio conceptual) que a menudo se encuentra en dimensiones similares de menor nivel, compartiendo una estructura discriminativa

común. Es allí donde el **TL** puede resultar beneficioso y ser aprovechado por **TDA**. La **figura 3.7** muestra cómo, a partir de cuatro clases de datos agrupados (1 y 2 etiquetadas, 3 y 4 sin etiquetar) se desea hallar el subespacio discriminativo adecuado y se dan tres situaciones, donde PCA+K-medias falla en cualquier caso, DisK-medias funciona cuando las clases están ligeramente escalonadas pero no así al estar en paralelo o muy ecalonadas; sin embargo, ayudado de clases etiquetadas encuentra satisfactoriamente el subespacio buscado.

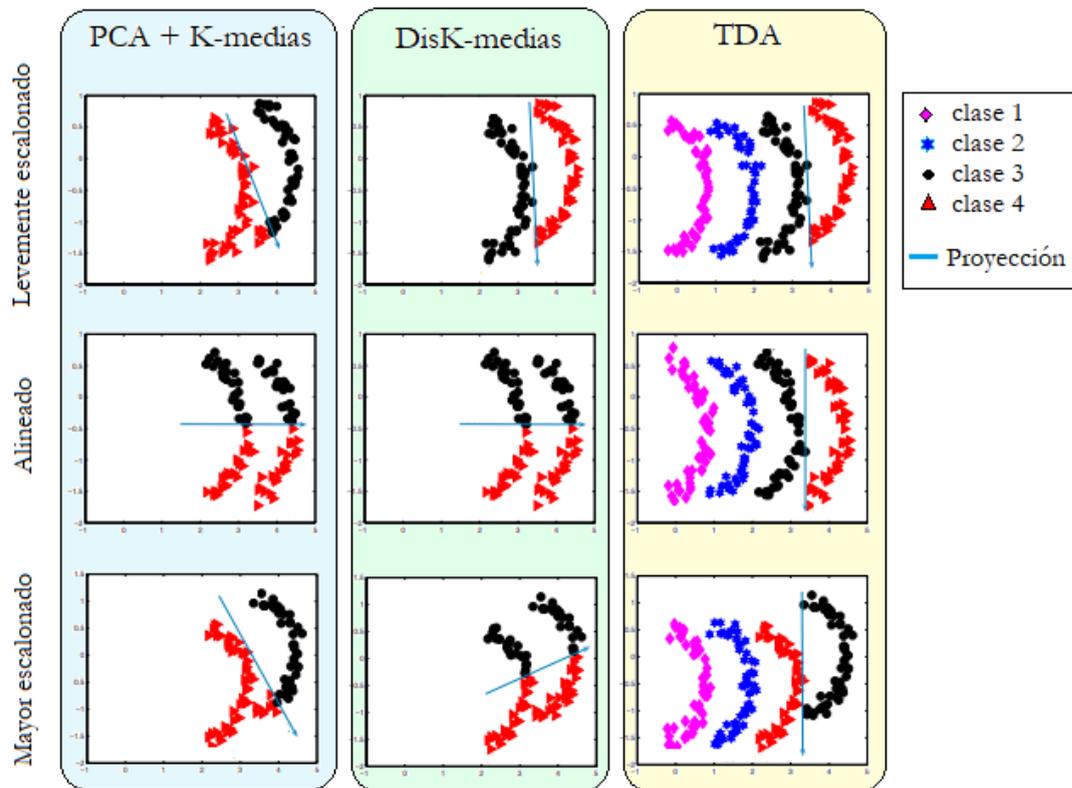


Figura 3.7: Desempeño de métodos PCA+K-medias, DisK-medias y TDA
 Figura adaptada de **Transferred Dimensionality Reduction** - Wang, Song y Zhang

Para evitar confusión, cabe destacar que si bien este trabajo mantiene puntos de similitud con **SsL**; en este contexto particular se aborda la resolución de una única tarea así como la transferencia de conocimiento desde información etiquetada hacia no etiquetada, admitiendo muestras etiquetadas de diferentes clases (con características subyacentes compartidas con la tarea objetivo) y perteneciendo dichos datos a diferentes tareas, mientras que en el SsL tradicional los datos etiquetados y sin etiquetar comparten dominios de clase y cada clase debe contar tanto con datos etiquetados como sin etiquetar con idéntica **distribución** de probabilidad.

El algoritmo **TDA** aborda el problema **TDR** mediante la extracción discriminativa de datos etiquetados y los transfiere a una reducción de dimensionalidad discriminativa para revisar iterativamente los resultados. Finalmente, utilizando tanto datos etiquetados como sin etiquetar de diferentes clases, se busca un subespacio discriminativo y de manera simultánea, un agrupamiento óptimo. Dado que los datos etiquetados y sin etiquetar provienen de diferentes conjuntos pero de un dominio conceptual común, sabemos que la estructura discriminativa puede compartirse, efectivizando la transferencia de información.

TDA consta de una fase de inicialización para el agrupamiento de los datos sin etiquetar utilizando **K-medias**; seguido de una fase de **reducción de dimensionalidad** donde se encuentra el subespacio óptimo; luego se efectúa una compactación de cúmulos para los datos objetivo donde las instancias sin etiquetar se agrupan en base al subespacio óptimo descubierto, hasta lograr convergencia y finalmente, se incrustan las muestras en el subespacio. Los experimentos han demostrado que el algoritmo converge con relativa rapidez (menos de 10 iteraciones) y su complejidad computacional está determinada por la dimensión de los datos originales, pudiéndose recurrir inicialmente **PCA** para su reducción y la aceleración de ejecución.

Como mencionamos en trabajos previos, una interrogante crucial a considerar es: “¿qué información transferir?”, usualmente existen varias clases de muestras etiquetadas en el conjunto de datos conocido, pero no todas son de ayuda para la tarea específica no supervisada ya que sólo algunas coinciden en su estructura discriminativa con las muestras sin etiquetar, y utilizar las restantes sería perjudicial. Además, utilizar todos los datos demandaría excesivo tiempo, en consecuencia se necesita seleccionar un subconjunto apropiado de datos en nuestra tarea de aprendizaje para maximizar la habilidad discriminativa sobre los datos destino.

Los resultados experimentales muestran que TDA supera notablemente el rendimiento sobre otros métodos no supervisados para un número de clases (**K**) sin etiquetar relativamente bajo, debido a que la estructura discriminante objetivo desconocida se complejiza al aumentar **K**, dificultando su adecuada descripción; mientras que el aumento de clases etiquetadas (**C**) no afecta el rendimiento (ver **tabla 3.6**).

Conjunto de datos	Métodos			
	PCA	LPP	DisKmedias	TDA
AT&T				
C=2, K=2	0.80 (0.15)	0.72 (0.12)	0.91 (0.16)	1.00 (0.02)
C=2, K=3	0.90 (0.11)	0.78 (0.08)	0.90 (0.13)	0.96 (0.10)
C=2, K=4	0.86 (0.10)	0.80 (0.08)	0.89 (0.12)	0.91 (0.11)
C=3, K=2	0.84 (0.13)	0.70 (0.13)	0.89 (0.16)	1.00 (0.02)
C=3, K=3	0.93 (0.08)	0.81 (0.08)	0.89 (0.15)	0.97 (0.07)
C=3, K=4	0.88 (0.08)	0.83 (0.06)	0.88 (0.10)	0.92 (0.10)
C=4, K=2	0.84 (0.14)	0.69 (0.11)	0.89 (0.20)	1.00 (0.02)
C=4, K=3	0.90 (0.11)	0.80 (0.10)	0.86 (0.14)	0.97 (0.07)
C=4, K=4	0.88 (0.08)	0.79 (0.09)	0.91 (0.11)	0.92 (0.10)

Tabla 3.6: Resultados sobre el conjunto de datos de AT&T con métrica ACC
 Tabla adaptada de **Transferred Dimensionality Reduction - Wang, Song y Zhang**

El aporte relevante de este trabajo gira en torno a la presentación el problema **TDR** y su resolución mediante un novedoso enfoque propuesto por el algoritmo **TDA** que aprovecha datos etiquetados y sin etiquetar de diferentes dominios, transfiriendo información discriminativa específica desde el conocimiento supervisado a muestras sin etiquetar sobre otros dominios de clase y descubre el subespacio más adecuado para la incrustación dimensional inferior, ésta situación es ampliamente aplicable sobre conjuntos de datos sin etiquetar que experimenten un drástico aumento, como sucede usualmente en internet. La experimentación desarrollada demuestra que la solución propuesta por TDA goza de eficiencia y robustez, adjudicándose como superador respecto de los mejores métodos no supervisados tradicionales.

3.10 - Transferencia entre dominios para el aprendizaje por refuerzo

Matthew Taylor y Peter Stone (2007) [10]

Este trabajo se enmarca dentro de la técnica de **RTL**, detallada en la **sección 2.6.4**, ya que tiene por objetivo acelerar el proceso de aprendizaje de un determinado agente objetivo. En el contexto del **RL**, la mayoría de los métodos introducidos transfieren el conocimiento entre tareas estrechamente relacionadas, con **agentes** desempeñándose en escenarios similares con repertorios de acciones y estructuras de recompensas semejantes. El aporte y valor de este trabajo reside en la demostración de la factibilidad de transferencia de conocimiento entre dominios y su efectividad cuando la tarea de origen se selecciona desde un dominio relativamente simple, siendo la tarea destino más compleja. Con ese fin se introduce el novedoso **método de transferencia de reglas (RTm)**, un algoritmo de transferencia de RL independiente del **dominio** que inicialmente aprende reglas para extraer una política a partir de la **tarea** origen (T_S), para luego aprovecharlas a fin de acelerar el aprendizaje sobre una **tarea** objetivo (T_T). Además de su desempeño de transferencia entre dominios, RT presenta bajos requisitos computacionales y admite la transferencia entre agentes con representaciones internas diferentes, posibilitando un mayor desacoplamiento entre ellos.

El funcionamiento de **RTm**, consiste de las siguientes etapas:

1. Aprender una política ($\pi: E \rightarrow A$; con E y A , conjuntos de estados y acciones) sobre T_S .
2. Aprender una lista de decisiones de origen ($D_O: E \rightarrow A$) que resume la política de origen.
3. Modificar la lista de decisiones aprendida (D_O) para utilizarla sobre T_T ($\text{Traducir}(D_O) \rightarrow D_D$). Si T_T contiene variables de estado y/o acciones diferentes a las de T_S , no es posible su aplicación directa, debiendo definir funciones de traducción o mapeo para utilizar D_D .
4. Utilizar la lista D_D para aprender una política sobre T_T .

Lo novedoso de este método es que aprovecha las reglas para brindar una representación abstracta de una política de T_S utilizable sobre T_T . La utilización de reglas es eficiente ya que son rápidamente aprendidas y legibles por personas; aunque se admiten otras representaciones internas, como las redes neuronales. La utilización de reglas como representaciones intermedias, posibilita el desacople de las técnicas particulares de aprendizaje de ambas tareas, abstrayendo el comportamiento del agente de origen para ser consolidado por el agente destino. Para que la transferencia de reglas resulte efectiva, se considera la lista de decisiones traducida D_D como recomendaciones más que como reglas a seguir, donde inicialmente el agente se beneficia de la lista de decisión pero va refinando su política al acumular experiencia sobre T_T .

Se definen tres métricas para la utilización de consejos:

1. **Bonificación de valor**, que calcula qué acción de la lista de acciones transferida (D_D), debería ser recomendada en base al estado actual y la bonifica favoreciendo su elección.
2. **Acción adicional**, que agrega una pseudo-acción a la descripción de acciones de T_T y al inicio del entrenamiento se fuerza su elección para guiar su utilización.
3. **Variable adicional**, que agrega una variable de estado a la descripción de estados de T_T , luego para conocer su importancia, fuerza al agente a utilizarla y a medida que el entrenamiento avanza, éste aprende a ignorarla en caso que su elección no sea óptima.

Se definen tres tareas de referencia para la experimentación pertenecientes al contexto de juegos:

1. **Mantente alejado (MAle)**, donde un equipo de 3 guardianes intentan mantener posesión de una pelota dentro de un área mientras que 2 ladrones trabajan para evitarlo. El aprendizaje en este dominio multi-agentes es complejo dadas las acciones estocásticas, el espacio de estados continuo y las observaciones ruidosas (simuladas) en los sensores de los agentes.

2. **Mundo anillo** (MAni), donde un jugador debe dirigirse a dos objetivos determinados en un anillo y evitar que un oponente lo intercepte. El jugador puede avanzar hacia algún objetivo o permanecer estático, mientras el oponente se aproxima incrementando la probabilidad de captura. No hay ruidos en las percepciones de agentes y el único factor estocástico es la aleatoriedad asociada con la probabilidad de captura.
3. **Justa de caballeros** (JdC), donde en un tablero cuadrado dividido en celdas, jugador y rival comienzan en lados opuestos alternando movimientos; el objetivo del jugador es llegar al extremo opuesto sin ser alcanzado por el oponente. No existe ruido en la percepción.

Las tareas de origen JdC y MAni se diseñaron con grados de similitud incremental, respecto de la tarea objetivo MAle. A fin de demostrar que la transferencia desde MAni puede mejorar significativamente el rendimiento considerando tres métricas de transferencia (rendimiento inicial, rendimiento asintótico y recompensa acumulada). Además, la transferencia desde JdC, al ser menos parecida a MAle por poseer menos variables de estado, función de transición menos similar y estructura de recompensas muy diferente, sólo puede incrementar significativamente el rendimiento sobre una de las métricas (rendimiento inicial). Esto también se vio reflejado en los resultados que expusieron que el agente destino siguió en mayor proporción los consejos generados desde MAni (90 %) que desde JdC (85 %), indicando que la utilidad de los consejos está directamente relacionada al grado de similitud entre dominios, la **figura 3.8** resume estos resultados y evidencia las mejoras producidas por el TL analizadas en el marco teórico de este trabajo (ver **sección 2.4**). También se analizó la sensibilidad de **RT** respecto a ciertos parámetros, pudiendo determinar que variar el diámetro del anillo en MAni impacta en su afinidad con MAle, permitiendo la parametrización óptima de T_s que maximice el rendimiento de transferencia.

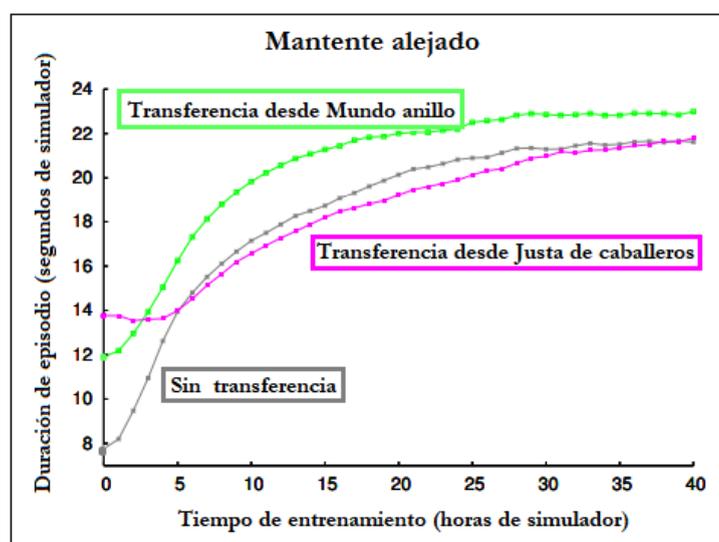


Figura 3.8: Curvas de transferencia de aprendizaje hacia MAle

Figura adaptada de **Cross-Domain Transfer for Reinforcement Learning** - Taylor y Stone

Los experimentos ratifican la flexibilidad de RT , así como la capacidad de los agentes para transferir exitosamente el conocimiento. En consecuencia, concluimos que la transferencia entre dominios optimiza el aprendizaje si se conocen las relaciones intervinientes. En síntesis, RT brinda una solución razonable al problema de mapeo cuando T_s se encuentra suficientemente comprendida, mientras que T_T es desconocida.

El abordaje de este trabajo ha evidenciado que gracias al análisis de datos y la evaluación intuitiva de T_T es factible construir tareas de origen cuya transferencia represente la mejora sustancial del aprendizaje.

3.11 - Distant Domain Transfer Learning

Ben Tan, Yu Zhang, Sinno Jialin Pan y Qiang Yang (2017) [33]

Este trabajo se enmarca dentro de la técnica de **transferencia transductiva (TTL)** donde las tareas involucradas en el proceso de transferencia coinciden pero los dominios considerados difieren (ver **tabla 2.1**), siendo el aspecto distintivo de este trabajo que dicha diferencia resulta considerable. Además, debido a que los espacios de características asociados pueden diferir, este es un caso de **TL heterogéneo (HetTL)**, donde el método de transferencia utilizado deberá adaptar no sólo las distribuciones sino también estos espacios.

El análisis de este trabajo resulta conveniente ya que se enfoca en la resolución de problemas de **transferencia de conocimiento entre dominios distantes (DDTL)** donde no se asume una correlación estrecha entre dominios de transferencia y se propone el **algoritmo de aprendizaje selectivo (SLA)** que intuitivamente utiliza un **autoencoder** convolucional supervisado para la selección gradual de datos de utilidad sin etiquetar desde dominios intermedios y los utiliza a modo de puente para restringir la brecha entre distribuciones de dominios y posibilitar la transferencia de conocimiento. Los resultados empíricos han demostrado el incremento del 17 % en la precisión de clasificación alcanzada por SLA.

La motivación para la resolución de problemas de este tipo se inspira en la transitividad y capacidad de inferencia del aprendizaje humano, es decir, la utilización de conceptos intermedios como puente para el aprendizaje de nuevos conocimientos entre conceptos aparentemente no relacionados. Se debe considerar que los algoritmos de TL no trabajan bien cuando los dominios están conceptualmente distantes o débilmente relacionados ya que no comparten características comunes de forma u otros aspectos. En esos contextos donde no existe una relación clara entre los dominios implicados, se debe tener especial cuidado en evitar que el enfoque resolutivo incurra en **NT**, degenerando en un rendimiento inferior al pretendido.

El método SLA tiene como componente principal un **autoencoder** y funciona seleccionando simultáneamente instancias útiles desde un dominio de origen (D_S) e intermedios (D_I), aprendiendo representaciones de alto nivel de los datos seleccionados y entrenando un clasificador sobre el dominio objetivo (D_T). El proceso de aprendizaje iterativo añade selectivamente nuevos puntos de datos desde D_I y remueve aquellos que resultan de poca ayuda en D_S evolucionando un modelo específico de D_T hasta satisfacer un determinado criterio de detención, como se aprecia en la **figura 3.9** para la transferencia entre tareas “rostro a avión” y “rostro a reloj” observando la utilización de diferentes instancias desde cada D_I .



Figura 3.9: Proceso de transferencia en SLA con dominios intermedios
Figura adaptada de **Distant Domain Transfer Learning - Tan, Zhang, Pan y Yang**

En base a los resultados experimentales alcanzados, este trabajo demuestra que el algoritmo SLA puede alcanzar un rendimiento asombroso en términos de precisión consolidándose como “estado del arte” en el área. Por consiguiente, uno de sus aportes más relevantes es justamente la posibilidad de utilización de instancias útiles desde un conjunto de dominios intermedios para efectivizar la transferencia; así como también considerar la aplicabilidad de técnicas de TL sobre tareas pertenecientes a dominios muy disímiles.

3.12 - Conclusión

En este capítulo, se expusieron algunas de las principales y más recientes metodologías resolutivas y trabajos de investigación en el área de **TL**. Así mismo, se ha pretendido exponer trabajos que aborden temáticas poco relacionadas, pertenecientes a diferentes subáreas previamente definidas (técnicas y enfoques), a fin de abarcar un amplio espectro de situaciones y contextos que resulten nutritivas para el lector.

Se pueden encontrar muchos escenarios propios de la ingeniería de conocimientos donde el TL puede resultar realmente beneficioso. A fin de ilustrar esos casos, se destina una sección del capítulo donde se presentan tres escenarios que han motivado el desarrollo del TL.

Siguiendo como principal referencia la división de **técnicas** y **enfoques** presentados en el **capítulo 2**, se aborda un trabajo relacionado en cada subsección. Esto, además de servir como guía orientativa al lector, permite individualizar diferentes contextos y vicisitudes que intenta afrontar y superar cada subárea.

El presente capítulo tuvo por objetivo la exposición de una serie de investigaciones y trabajos de relevancia que han tomado carácter de referenciales en cada subárea exponiendo diferentes abordajes. Su lectura tiene por objetivo proponer al lector la posibilidad de contemplar distintas herramientas disponibles, delinear las áreas de investigación actuales dentro del TL y vislumbrar su aplicabilidad en cada problemática del mundo real que se pretende resolver.

Capítulo 4 - Propuesta metodológica para la aplicación de TL

4.1 - Introducción

En este capítulo se describirá una propuesta de alto nivel de abstracción y generalidad, con el objetivo de aumentar su aplicabilidad sobre un variado espectro de problemáticas del mundo actual, plausibles de ser abordadas y resueltas mediante ML. Se procederá a estudiar y definir las etapas necesarias que deben concretarse para la resolución de un determinado problema utilizando técnicas de ML y para tal fin, este capítulo se organiza bajo una estructura de fases o etapas que posteriormente se profundizarán y extenderán mediante la presentación de un conjunto de subtarefas que las componen.

Este capítulo tiene por objetivo esquematizar un flujo de pasos donde se evidencien los distintos procesos que deben cumplimentarse cuando se desea solventar ciertas dificultades que pueden acontecer en el campo del ML y que pudieran dificultar la obtención de modelos robustos de elevado rendimiento. De esta forma se consolidará en un flujo de procesos lo analizado en los capítulos previos, permitiendo que el lector de la tesina (ya sea estudiante, investigador o profesional de IT) que desee implantar este flujo resolutivo sobre cualquier problemática concreta, pueda comprender cada etapa y alcanzar las metas propuestas.

Se pretende que el actual capítulo haga las veces de nexo entre los contenidos vistos previamente y su consolidación en lineamientos resolutivos aplicable a gran cantidad de problemáticas que puedan suscitarse en el mundo actual y que además sirvan de fundamento para la aplicación concreta. El mismo representa la continuidad de los conocimientos teóricos hacia un plano más pragmático donde se volcará lo aprendido en fundamentos y trabajos relacionados, hacia la descripción de un caso aplicativo práctico y representativo.

4.2 - Etapas del proceso resolutivo

Para el análisis, desarrollo y resolución de cualquier problemática mediante la utilización de técnicas de ML es necesario delinear y atravesar una serie de procesos que necesariamente deben ser definidos de forma adecuada y precisa, a fin de completar dicho proceso resolutivo complejo como un recorrido donde ningún tópico fundamental deje de ser considerado y abordado.

Por lo expresado anteriormente se procederá, en esta sección, a la enumeración de una serie de etapas consideradas como trascendentales para el abordaje, desarrollo y logro de los objetivos planteados al abordar cualquier proyecto de ML, dentro de cada etapa se describirán las posibles dificultades que puedan experimentarse así como las decisiones a evaluar y tomar para superarlas. Finalmente, se intentará clarificar aún más dicha propuesta metodológica plasmando las etapas descritas sobre un diagrama para brindarle al lector una visión simplificada y concisa de la serie de actividades a cumplimentar. Resulta pertinente en este punto, aclarar que no se hará una exhaustiva profundización de cada una de las etapas y subtarefas asociadas ya que este trabajo se dejará a cargo de la próxima sección; representando la profundización y refinamiento de los tópicos de este capítulo, un posible tema de desarrollo de futuras tesinas.

Las etapas fundamentales que cualquier proceso resolutivo debe considerar son las siguientes:

1. Identificación, definición y comprensión de la problemática

Etapa fundamental del proceso resolutivo, donde debe establecerse claramente la temática objetivo que el proyecto abordará, escudriñando su contexto, considerando las implicancias relacionadas (quiénes y de qué modo se beneficiarán o de qué modo los impactará), analizar el flujo de trabajo usual asociado y el grado de masividad de su utilización, aplicativos software existentes y hardware demandado, así como cualquier factor que contribuya a la comprensión de la problemática a tratar.

2. Estudio de la problemática en el contexto del ML

En esta etapa se debe proceder al análisis de todo lo relativo al problema actual bajo marco contextual del aprendizaje automático, estableciendo cuál es el paradigma resolutivo más adecuado para el tipo de problema particular (el determinado en la etapa previa), analizar los datos requeridos, considerar trabajos previos o algoritmos propuestos en este campo de interés, a fin de aprovechar al máximo todo el conocimiento posible para evitar errores comunes y seleccionar los caminos más fructíferos a partir de la experiencia disponible recabada desde fuentes confiables. También se debe proceder al análisis y enumeración de las premisas, así como el grado de flexibilidad inherente que cada una de ellas presenta para el proyecto.

3. Preparación de los datos

Etapa de destacada importancia para el abordaje de cualquier problema de ML y que conlleva alrededor de la mitad del tiempo empleado en un proyecto. En ella debe estudiarse con detenimiento, qué conjuntos de datos se encuentran disponibles para la generación del modelo deseado, analizar factores como cantidad de instancias, atributos de los cuales se compone, dimensionalidad, accesibilidad, gratuidad de los mismos, así como también contemplar y evaluar la aplicación de técnicas de generación y aumentación disponibles sobre los datos.

Al fin y al cabo, y conociendo que un mayor rendimiento del modelo de ML construido será asequible a mayor cantidad de instancias disponibles y mayor grado de representatividad de las mismas, se necesita poder responder cuestiones relativas a la cantidad de instancias utilizables (en etapas de entrenamiento y validación), estructura, organización y esquema de particionamiento de datos, metodologías, métricas y parametrizaciones más adecuadas para cada etapa, entre otras.

4. Construcción del modelo

En base al paradigma resolutivo actual (definido utilizando el conocimiento contenido en el **capítulo 2**, en conjunción con lo analizado en la **etapa 2**), la fisonomía del conjunto de datos (analizada en la **etapa 3**) y las metodologías de ML más apropiadas (de forma similar al análisis del **capítulo 3** para TL) para el abordaje del problema actual (identificado en la **etapa 1**); se procede a la generación del modelo de ML, consistente del entrenamiento, evaluación y parametrización de forma iterativa (en épocas) y el almacenamiento de los resultados obtenidos para su seguimiento y correcta trazabilidad a futuro, que serán la base para subsiguientes estudios comparativos y mensurar apropiadamente los avances hechos en próximos trabajos.

5. Evaluación del modelo

A partir de los datos recolectados en la etapa previa y en añadidura con los conocimientos incorporados en el recorrido de cada etapa descripta, se procede al análisis comparativo y el contraste de los resultados obtenidos contra los establecidos en las premisas u objetivos explícitos a cumplimentar por parte la solución implantada (definidas en la **etapa 2**). En caso que las mismas no se hayan podido satisfacer completamente o aún satisfechas se considere la implantación de soluciones superadoras; toma protagonismo el estudio y aprovechamiento de técnicas de aprendizaje por transferencia.

6. Consideración e implantación de técnicas de TL

Esta etapa sintetiza el fundamento y propósito del presente trabajo, ya que es en este punto donde lo expuesto en capítulos previos debería bastar como argumentación a fin que el lector comprenda a cabalidad que la aplicación de técnicas de TL tendrán el potencial de incrementar de manera significativa la calidad del modelo construido y la satisfacción de los usuarios.

Requiere el análisis cuidadoso de las diferentes técnicas y enfoques de TL en base al caso de aplicación concreto (como grado de relación entre tareas, disponibilidad de datos etiquetados, cantidad de instancias del conjunto de datos, factores de dinamismo, etc.), así como la revisión de las metodologías exitosas bajo el contexto y paradigma resolutivo considerado a fin de asegurarnos que su empleo se adecúa al escenario objetivo; además deberían contemplarse los factores que pudieran ejercer como limitantes o degradantes del rendimiento; consolidándose en esta fase, los contenidos vistos en los capítulos 2 y 3 de esta tesina.

7. Control de resultados

Etapa considerada de vital importancia debido a que en ella se contrastan los resultados obtenidos por los modelos que han sido generados bajo el paradigma de ML tradicional o haciendo uso del TL. El objetivo es conocer y ponderar de forma precisa, el cumplimiento de las premisas planteadas, así como el grado de mejora alcanzado mediante la aplicación de técnicas de TL. En casos donde se incumplan premisas esenciales utilizando ML tradicional, no se hayan observado mejoras de rendimiento o incluso evidenciarse su deterioro (**NT**) habiendo aplicado TL, debería procederse a la reconsideración de las técnicas y enfoques aplicados.

8. Despliegue

Etapa final, pero no por eso de menor importancia, consistente en el desarrollo de los factores relativos a la puesta en producción del modelo de ML refinado, robusto y de elevada calidad, sobre los dispositivos o aplicaciones que usufructuarán su poder predictivo (sitios webs de negocios o didácticos, dispositivos móviles, de audio o video, de seguridad, artefactos IoT, automóviles, robots, equipos de diagnóstico por imágenes, etc.) y el desarrollo de los procesos de optimización necesarios para su apropiado funcionamiento sobre su entorno productivo.

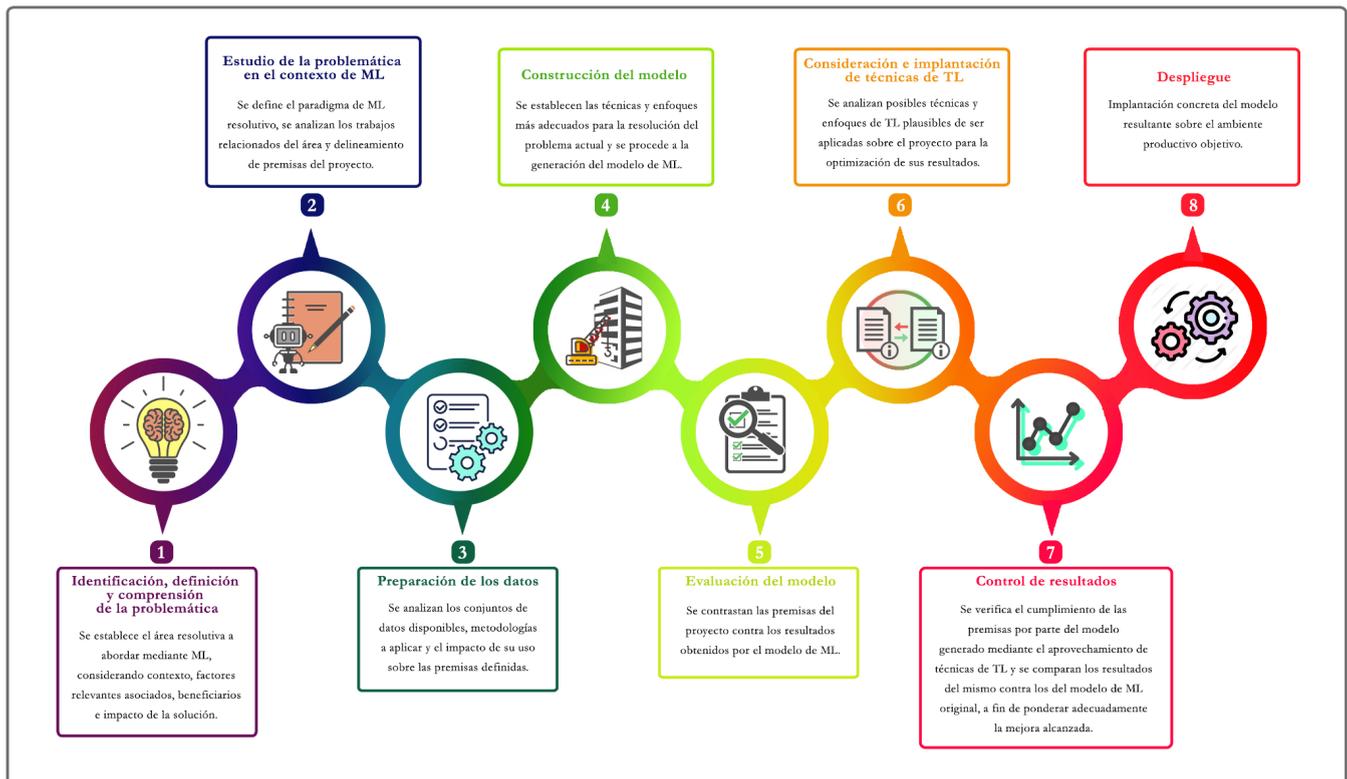


Figura 4.1: Etapas resolutivas en un proyecto de ML
Figura de elaboración propia

4.3 - Flujo metodológico

Esta sección se aboca a realizar la ampliación de cada una de las fases descritas en la sección previa, apuntando a su profundización mediante la descripción de un conjunto de subtarefas o procesos que los componen; por consiguiente se hará foco en la descripción de cada una de ellas detallando los trabajos asociados y su ubicación dentro de las etapas vistas, propias de cualquier proyecto de ML. De este modo, se pretende que el lector pueda disponer de un mayor grado de especificidad y detalle de los conocimientos presentados, a fin de evitar pasar por alto alguna y disponer de un conjunto de secuencias de acciones concretas y verificables que realizar, para la obtención de resultados óptimos o bien cuando los mismos no alcancen el nivel de satisfactoriedad deseado. De esta forma, se ponen de manifiesto las labores que cualquier investigador debe efectuar y que la mayoría de los proyectos de ML extendidos con TL exigen.

La etapa 1, correspondiente a la **identificación, definición y comprensión de la problemática**, puede subdividirse en las siguientes tareas:

- Establecimiento del área resolutive, donde se debe identificar de manera precisa y acotada el área sobre la cuál se desea llevar a cabo la solución u optimización. Algunas de las áreas más importantes son: análisis del habla, diagnóstico médico o cuidado de la salud, prevención de fraudes informáticos, robótica y visión por computadora, entre tantas otras.
- Comprensión del contexto aplicativo, aquí debe considerarse el ambiente donde se ejecutará la solución propuesta y diferentes factores que determinan restricciones sobre dicha solución.

- Utilidad y población beneficiaria, es fundamental identificar a la población objetivo que utilizará el modelo de ML ya que a mayor grado de utilidad e impacto (cantidad de personas beneficiadas), mayor será el atractivo y el valor asociado del modelo resultante.
- Aplicaciones existentes, una vez determinado el problema objetivo resulta de mucha utilidad indagar en las resoluciones similares o relacionadas en dicho tópico, analizando ventajas y dificultades de cada una, evolución histórica y proyección futura.

La etapa 2, correspondiente al **estudio de la problemática en el contexto del ML**, puede subdividirse en las siguientes tareas:

- Establecimiento del paradigma resolutivo, donde el desarrollador debe analizar las opciones disponibles y seleccionar el paradigma de ML adecuado según el problema abordado, algunos son: supervisado (**SLP**), semi-supervisado (**SsLP**), no supervisado (**ULP**) y reforzado (**RLP**).
- Investigación de trabajos relacionados, una vez identificado el paradigma sobre el cuál se sustentará la solución implementada, debería procederse a la investigación y análisis cuidadoso de los trabajos existentes bajo esos mismos paradigmas y relacionados al contexto actual.
- Definición de premisas del proyecto, aquí se describen una serie de requisitos a cumplimentar por parte de la solución implementada, algunas de las premisas del proyecto tendrán carácter de inamovibles, siendo su cumplimiento obligatorio, algunas de cumplimiento más flexible o parcial y otras deseadas pero de cumplimiento opcional. Girando en torno a factores como:
 - Precisión, se establece un rango de precisión aceptable para el modelo de ML final.
 - Utilización de recursos, se definen los requerimientos para la generación del modelo buscado, pueden ser tanto de hardware, software, de datos, recursos humanos, conocimiento especializado, tiempos, etc.
 - Tiempo requerido por cada etapa generativa, se establece el tiempo que debería ocupar cada etapa constructiva del modelo (entrenamiento y validación).
 - Hardware requerido para el entrenamiento, se determina sobre qué plataforma deberá ser entrenado el modelo (CPUs, GPUs, cloud computing, etc.).
 - Conjunto de datos necesario, se evalúan cuidadosamente los diferentes conjuntos de datos que pueden utilizarse e involucra factores como:
 - Accesibilidad, la facilidad de acceso a los datos es un factor importante.
 - Gratuidad, impacta sobre la implantación de la resolución, su costo y facilidad de extensión futura.
 - Disponibilidad y continuidad evolutiva, debe contemplarse para futuras aplicaciones compatibles, optimizaciones o extensiones.
 - Cantidad de instancias, dato fundamental ya que seguramente impactará en la precisión del modelo final y el cumplimiento de las premisas.
 - Dimensionalidad, se debe contemplar cuidadosamente y conjuntamente con la cantidad de ejemplos, ya que impacta en la calidad del modelo.

La etapa 3, correspondiente a la **preparación de los datos**, puede subdividirse en las siguientes tareas:

- Análisis y estudio de datos disponibles, ya determinado el conjunto de datos a utilizar es fundamental revisar cuidadosamente cada uno de los atributos de los cuales se compone y realizar estadísticas sobre ellos que puedan indicar tratamientos previos para su refinamiento. En esta etapa tenemos: el análisis exploratorio (**EDA**), generación de datos estadísticos, visualización y análisis relacional entre variables, minería y extracción de datos.
- Adquisición y almacenamiento, selección concreta del conjunto de datos a utilizar, almacenamiento de datos “crudos” (sin refinar o preprocesar), limpieza de datos erróneos, imputación de datos faltantes e incorporación de información (ingeniería de características).
- Implementación de técnicas de aumentación, ante un conjunto de datos muy limitado en cantidad de ejemplos, es posible incrementarlo considerablemente mediante la incorporación de datos generados artificialmente. Se dedica el **anexo A** en este trabajo referido a ese tema.
- Aplicación de técnicas de reducción de dimensionalidad, determinar cuales de estas técnicas son aplicables en base a la fisonomía del conjunto de datos analizada previamente. Algunas de ellas requieren la incorporación de metodologías de tratamiento previo, como por ejemplo normalización. Se dedica el **anexo B** en este trabajo referido a ese tema.
- Análisis de segmentación de datos, buscando la división más eficiente y adecuada de los datos en las distintas etapas generativas (entrenamiento, validación y prueba).
- Evaluación de costos y cumplimiento de premisas, tarea final de esta etapa que tiene una importancia notable ya que consiste en establecer si los costos implicados en las etapas a cumplimentar sobre los datos no impiden el cumplimiento de premisas establecidas.

La etapa 4, correspondiente a la **construcción del modelo**, puede subdividirse en las siguientes tareas:

- Análisis de las técnicas más apropiadas bajo el paradigma elegido, analizando los pormenores implicados en la utilización de cada una de ellas y considerando conjuntamente la fisonomía de los datos (un trabajo muy similar al realizado en el **capítulo 3** de este trabajo, sólo que aquí se realiza sobre una metodología de ML tradicional determinada).
- Definición de algoritmos, métricas y parametrizaciones, en este paso y como consecuencia del paso previo se definen las más adecuadas metodologías y configuraciones del modelo objetivo.
 - Metodologías a utilizar, se establece el algoritmo concreto y sus parametrizaciones.
 - Etapa de entrenamiento, se define el proceso y la cantidad de iteraciones.
 - Determinación del sistema de validación, se determina la segmentación de datos en cada etapa, así como la metodología de evaluación a utilizar (simple, cruzada, Bootstrapping, LOOCV, comparación, etc.).

- Selección de métricas a utilizar (RMSE, MAE, AUC, Recall, etc.), establecidas como criterios de evaluación y seguimiento luego de las etapas generativas (entrenamiento y evaluación) analizando su proyección, detectando factores indeseados (como subajuste o sobreajuste) y concretando optimizaciones.
- Criterios de implantación, se determina si la generación se realizará utilizando recursos locales o disponibles en la nube (cloud), es relevante ya que los factores de utilización de recursos previamente analizados dependen en gran parte de este punto.
- Entrenamiento, proceso iterativo de evaluación del modelo y ajuste de parámetros, representa la implementación concreta de las etapas generativas y de aplicación las metodologías, métricas y parametrizaciones definidas anteriormente.

La etapa 5, correspondiente a la **evaluación del modelo**, puede subdividirse en las siguientes tareas:

- Análisis comparativo entre los resultados obtenidos y las premisas definidas:
 - Análisis de métricas de rendimiento y error establecidas, así como su progresión a lo largo de las etapas constructivas experimentadas.
 - Análisis de congruencia entre los resultados alcanzados por otros trabajos bajo idénticos paradigmas, técnicas y metodologías resolutivas y los obtenidos por el modelo generado.
 - Contraste de información, considerando los resultados evidenciados por el modelo de ML resultante y comparándolos contra las premisas definidas (precisión, error, tiempos, utilización de recursos, etc.).
- Retroalimentación, ajuste de hiper-parámetros del modelo, proceso de refinamiento iterativo para el reentrenamiento y mejora en las métricas de evaluación del modelo de ML resultante.

La etapa 6, correspondiente a la **consideración e implantación de técnicas de TL**, se subdivide en las siguientes tareas:

- Análisis de técnicas y enfoques adecuados, donde se investigan técnicas plausibles de aplicación y evalúan los resultados mostrados en áreas relacionadas a la problemática objetivo.
- Búsqueda de conjuntos de datos para el entrenamiento del modelo, a partir de las posibilidades de utilización de datos ampliadas que surgen de la utilización de técnicas de TL.
- Utilización de modelos pre-entrenados, presentes y disponibles para su uso por parte de la mayoría de los marcos de trabajo de ML (como PyTorch, Keras o TensorFlow).
- Análisis del grado de relación entre dominios de datos, debe estudiarse el grado de similitud relacional entre datos pertenecientes a ambos dominios (origen y objetivo) intervinientes en la transferencia; ya que la precisión del modelo obtenido, dependerá en gran parte de este factor.

La etapa 7, correspondiente al **control de resultados**, puede subdividirse en las siguientes tareas:

- Generación de información útil y análisis de resultados, se procede a recopilar, analizar y visualizar apropiadamente el rendimiento mostrado por el modelo TL (trazabilidad) y al análisis de los factores de incremento del rendimiento posibilitado por la incorporación de técnicas de transferencia frente al uso de ML tradicional.
 - Control de no ocurrencia de transferencia negativa (NT), donde los conocimientos vistos en la **sección 2.8** de este trabajo, cobran suma importancia.
- Control empírico del rendimiento, consiste en la revisión de los valores de mejora prometidos por la técnica y enfoque de TL utilizado, contra los empíricamente evidenciados por el modelo de TL, para que en caso de no ser congruentes se revea y solvente la causa del problema.
- Evaluación del cumplimiento de las premisas, se procede al control del cumplimiento de las premisas establecidas por parte del modelo generado mediante la utilización de TL.

La etapa 8, correspondiente al **despliegue**, puede subdividirse en las siguientes tareas:

- Migración del ambiente, una vez refinado nuestro modelo y sabiendo que es robusto, se procede a su despliegue sobre los dispositivos desde los cuales los usuarios lo utilizarán, por ejemplo para el despliegue sobre dispositivos móviles puede hacerse uso de aplicativos como CoreML, PyTorch Mobile o TensorFlow Lite, entre otros.
- Despliegue del modelo, implica la integración del modelo generado sobre el ecosistema aplicativo o herramienta, donde el mismo es utilizado mediante una interfaz de usuario (UI).

La **figura 4.2** mostrada a continuación, resume todos los procesos vistos dentro de un diagrama de flujo de actividades, donde se muestra la secuencia y orden de precedencia en el cual deben ser llevadas a cabo, así como los datos de entrada requeridos y los resultados que las mismas producen. Naturalmente, dicho ordenamiento resulta más bien orientativo ya que en casos prácticos, la secuenciación concreta de etapas, las posibles iteraciones entre diferentes actividades, así como también los flujos presentes en el diagrama, pueden ser mucho más complejos, rigiéndose por las particularidades de cada ámbito de aplicación. En otras palabras, este diagrama pretende ilustrar de forma general las tareas y flujos intervinientes en el proceso resolutivo, considerando que las variaciones posibles pueden ser tan amplias como los diferentes escenarios que puedan presentarse.

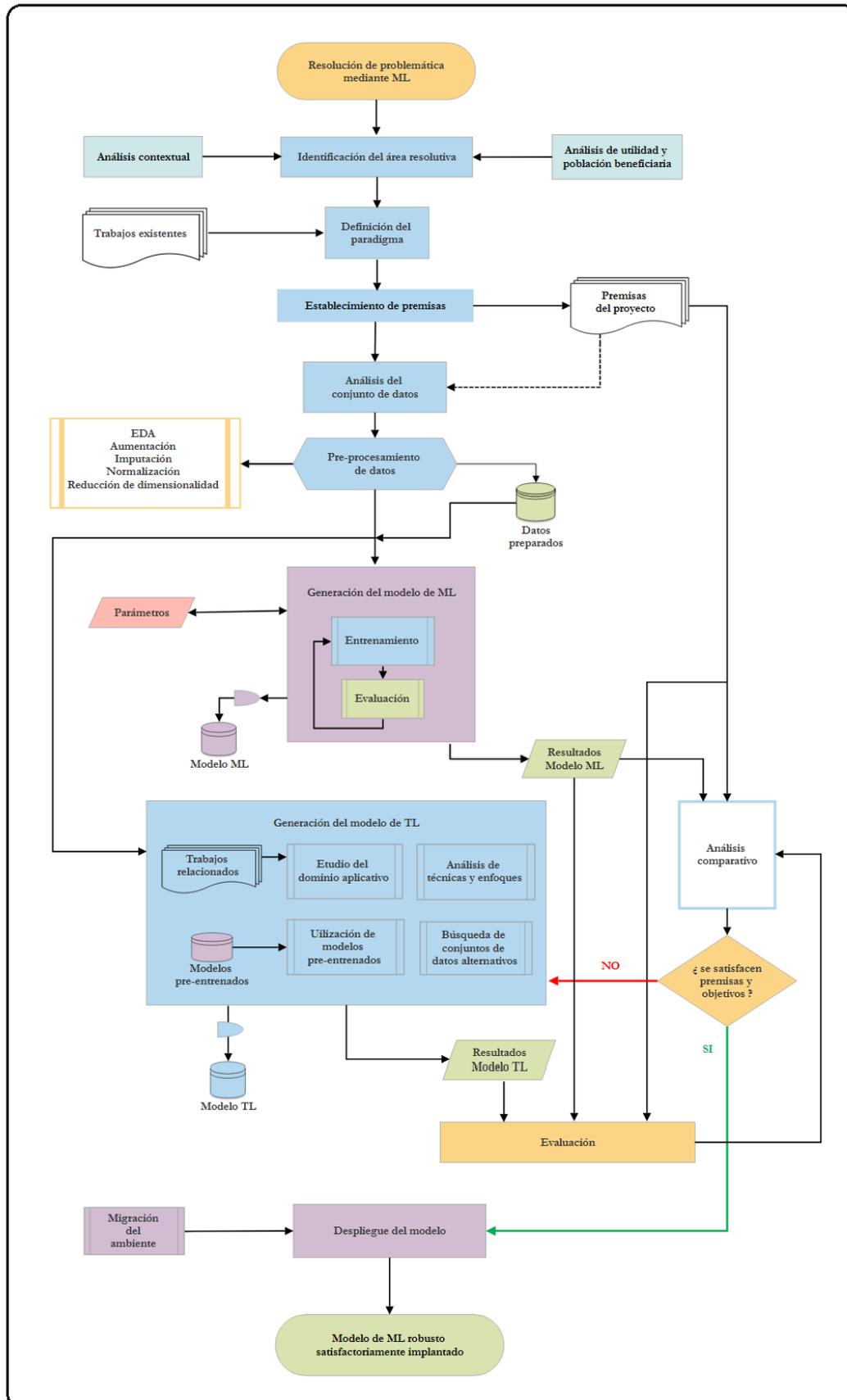


Figura 4.2: Diagrama de flujo de procesos en un proyecto de TL
 Figura de elaboración propia

Alcanzado este punto en el capítulo, resulta necesaria y oportuna la siguiente aclaración; si bien se ha planteado un esquema relativamente secuencial de etapas resolutorias dentro de un proyecto de ML, este escenario no siempre se presenta de manera tan lineal, en el mundo real. De hecho, y a modo de ejemplo, el análisis comparativo del cumplimiento de las premisas establecidas puede tener un rol mucho más activo entre etapas, pudiendo llegar incluso a evitar la implementación de algunas. Por ejemplo, en caso de incumplirse las premisas y saber que no es posible generar un modelo de ML robusto, resultará viable intentar solventarlo directamente mediante la aplicación de técnicas de TL sin la necesidad de desarrollar un modelo de ML tradicional. En el siguiente capítulo se verá un caso aplicativo donde precisamente acontece esta situación.

4.3.1 - Profundización de etapa de consideración e implantación de técnicas de TL

Como se ha descrito y graficado previamente en esta sección, la sexta etapa descrita correspondiente a la **consideración e implantación de técnicas de TL**, cuenta entre sus subtarefas más importantes de un proceso referido a la búsqueda, análisis y revisión de técnicas y enfoques más apropiado, dependientes del caso de aplicación puntual que se este abordando. Este proceso reviste cierta complejidad, representando además el corazón de esta tesina, motivo por el cuál se decidió la redacción de la actual subsección, con el objetivo de focalizar sobre dicha etapa y presentar una serie de reglas útiles aplicables por cualquier desarrollador o investigador, a fin de abreviar y simplificar el proceso de selección de metodologías de TL.

Tomando como punto de partida el grado de relación evidenciado por las tareas y dominios involucrados en el proceso de transferencia, es posible determinar bajo que técnica o enfoque resolutorio de TL situarnos según cada caso particular. En consecuencia, a partir de lo visto en la **sección 2.6** y subsiguientes pudimos determinar que existen diferentes dimensiones o aspectos distintivos bien diferenciados, que admiten ser considerados con el objetivo de clasificar los distintos escenarios para el abordaje de diferentes problemáticas y enlazarlos utilizando su relación con las distintas metodologías.

Entre las dimensiones distintivas que podemos mencionar, se encuentran:

- **Grado de relación entre tareas y dominios**

En base a esta dimensión distintiva (y conforme a lo mostrado en la **tabla 2.1**), sabemos que si ambos dominios coinciden pero las tareas difieren, deberían aplicarse técnicas de transferencia inductiva (**ITL**); en caso que ambos dominios difieran pero las tareas coincidan, aplicaríamos técnicas transductivas (**TTL**) y finalmente ante situaciones donde dominios y tareas difieran, deberían considerarse técnicas no supervisadas (**UTL**) y reforzadas (**RTL**).

- **Disponibilidad o ausencia de datos etiquetados**

En base a esta perspectiva (y conforme a lo mostrado en la **tabla 2.2**) sabemos que en situaciones donde se dispone de datos etiquetados en el dominio objetivo (D_T), deberían indagarse y aplicarse técnicas de **ITL**; en caso de no disponer de datos etiquetados sobre D_T pero sí disponer de algunos sobre el dominio de origen (D_S), deberían considerarse técnicas de **TTL** y finalmente si no se dispusiera de datos etiquetados sobre ningún dominio involucrado, deberían implantarse técnicas de transferencia asociadas a **UTL** y **RTL**.

- **Grado de discrepancia entre dominios**

En base a esta dimensión distintiva podemos determinar si es necesario llevar a cabo un proceso de adaptación de los espacios de características propios de los dominios involucrados en la transferencia, es decir, en situaciones donde no coincidan los espacios de características (X_S y X_T o bien Y_S y Y_T) de los dominios intervinientes, deberían contemplarse técnicas de transferencia heterogénea (**HetTL**); mientras que en aquellas ocasiones donde los espacios de características coincidan y solamente se requiera un proceso de adaptación sobre distribuciones de probabilidad asociadas, deberían aplicarse técnicas de transferencia homogénea (**HomTL**).

- **Situación de características y parámetros de dominios**

En base a esta dimensión discriminativa (y conforme a lo mostrado en la **tabla 2.3**), podemos determinar cuál de los enfoques disponibles será el más adecuado dependiendo del análisis de los datos correspondientes a los dominios intervinientes en el TL.

De esta forma, cuando sólo ciertas partes de los datos de D_S pueden utilizarse con ayuda de algunos datos etiquetados presentes en D_T o cuando no existan datos etiquetados sobre D_T pero las distribuciones de probabilidad de ambos dominios coinciden, deberían investigarse enfoques asociados a la **transferencia de instancias**, tanto con técnicas de **ITL** como de **TTL**.

Cuando es posible utilizar datos sin etiquetar en D_T intentando minimizar las diferencias entre los dominios involucrados y es factible utilizar gran cantidad de datos en D_S , sin importar si están etiquetados o no, deberían abordarse enfoques relacionados a la **transferencia de representación de características**, dentro de técnicas de **ITL**, **TTL** o **UTL**.

En aquellas situaciones donde las tareas involucradas se encuentren relacionadas y compartan parámetros y distribuciones previas, es posible utilizar enfoques asociados a la **transferencia de parámetros**, bajo la técnica de transferencia inductiva (**ITL**).

Finalmente, en situaciones donde los dominios involucrados en la transferencia sean coincidentes y exista un fuerte grado de relación entre ellos, puede utilizarse el enfoque de **transferencia de conocimiento relacional**, sobre técnicas inductivas (**ITL**).

Con lo expuesto previamente, se pretende brindar al lector un mayor grado de profundización asociado a las herramientas analíticas y técnicas de TL disponibles, que puedan ser esgrimidas al momento de enfrentar cualquier proyecto realista en este contexto. El objetivo de esta subsección es servir como nexo lógico entre los conceptos expuestos en el **capítulo 2**, que representaron el marco teórico del trabajo, y las metodologías prácticas aplicables sobre una amplia gama de situaciones, como las analizadas en el **capítulo 3**; otorgado una perspectiva más abarcativa y complementaria que amalgame los tópicos allí propuestos.

A partir de la aplicación de las metodologías vistas en la presente subsección, el lector dispondrá de una guía analítica que le permitirá desembarcar en las técnicas y enfoques más apropiados, el análisis y tratamiento de un conjunto acotado de metodologías, focalizando en la investigación de los trabajos más pertinentes y relevantes a determinado contexto, salvando de esta forma, tiempo y esfuerzo significativo considerando la amplitud de ramificaciones que esgrime el campo del TL.

Para finalizar, y a modo de consolidación de los contenidos previos, se procederá a la realización de un flujo de tareas dedicado exclusivamente a la subtarea “Análisis de técnicas y enfoques”, perteneciente al proceso de “Generación del modelo de TL” mostrado en la **figura 4.2**, incluyéndose el flujo de procesos mostrado en la **figura 4.3**, completamente dentro de esa subtarea y representando su expansión funcional. De esta forma, se clarifica, extiende y completa adecuadamente el flujo de tareas descrito la sección previa, respetando un enfoque descendente que comenzó con el estudio general de las etapas propias de un proyecto abordado mediante ML y desembocó en un conjunto de dimensiones distintivas que orientan hacia las metodologías más apropiadas dentro del recorrido resolutivo en el campo del TL.

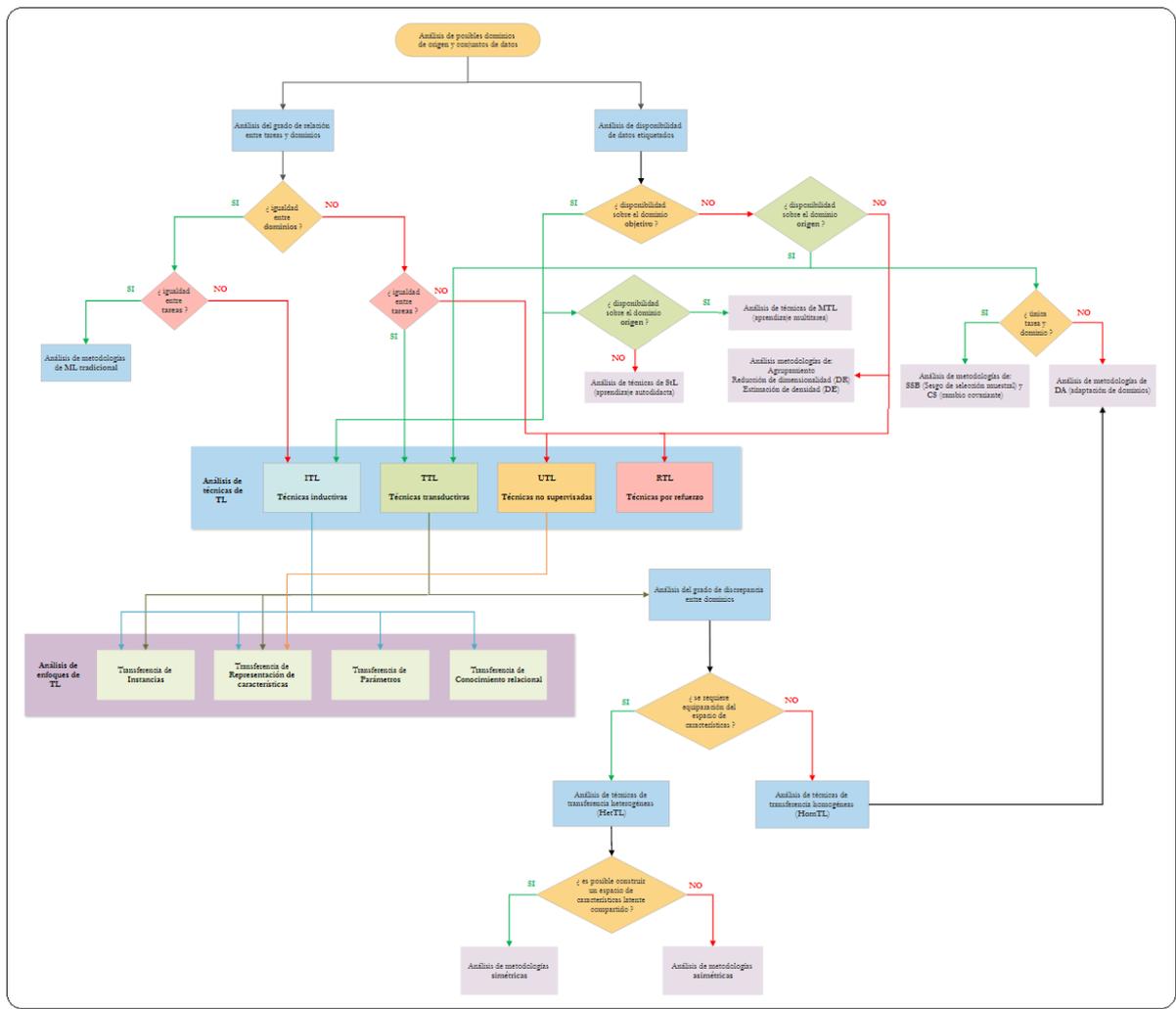


Figura 4.3: Diagrama de flujo de procesos del “Análisis de técnicas y enfoques” de TL
 Figura de elaboración propia

4.4 - Conclusión

En este capítulo se procedió a describir una propuesta esquemática de carácter y aplicación general que sirva como referencia útil para el lector acerca de los pasos que deben concretarse para lograr el desarrollo y

puesta en marcha de un proyecto de ML utilizando técnicas de TL con el propósito de superar las dificultades que puedan presentarse a lo largo del mismo.

Se realizó una categorización en etapas que describen el proceso de resolución mediante TL, desde el comienzo del análisis de la temática a resolver, hasta la evaluación de las ventajas obtenidas gracias a la aplicación de transferencia del conocimiento. Además se detallaron las tareas involucradas en cada etapa, plasmando finalmente dicho análisis en un diagrama de flujo, a fin de clarificar la metodología propuesta y con el objetivo de que pueda ser tomada como guía orientativa por parte del lector que afronte un trayecto resolutivo semejante. Finalmente se ahondó aún más en el proceso referente al análisis de técnicas y enfoques de TL, que representa el tópico central de este trabajo y la fusión de multitud de conceptos vistos.

El desarrollo del presente capítulo es apropiado y de relevancia ya que representa el nexo entre conceptos analizados previamente en capítulos que establecen el marco teórico (**capítulo 2 - Aprendizaje por transferencia**) y los principales trabajos recorridos en el área (**capítulo 3 - Estado del arte**); enlazando lo aprendido allí para proyectar dichos conocimientos sobre un caso de estudio (**capítulo 5 - Un caso aplicativo**) que represente la aplicación concreta de lo esquematizado en el presente capítulo. De esta forma, la redacción de este capítulo representa la continuidad lógica de los conceptos ya expuestos para avanzar desde la exposición de una variedad de conocimientos teóricos fundamentales, hacia un plano aplicativo. Este trayecto continuará marcándose en próximos capítulos, mediante la aplicación concreta de la metodología descrita aquí y el análisis de diferentes casos aplicativos actuales en el área del TL (**anexo C**).

El objetivo de este capítulo es amalgamar los conceptos y desarrollos asociados al TL vistos en capítulos previos y plasmarlos en una metodología resolutiva general aplicable a la mayor cantidad de problemas del área, lo que le otorgará gran utilidad pragmática. El lector puede, mediante la lectura del presente capítulo disponer de una guía orientativa que seguir, donde se expresen los lineamientos generales para la obtención de modelos de alta calidad utilizando TL.

Capítulo 5 - Un caso aplicativo

5.1 - Introducción

En este capítulo se presentará un caso de estudio aplicado, considerado como apropiado, relevante y representativo de los casos de aplicación de TL estudiados y que además posibilitará detallar el recorrido de aprendizaje que lo configura como el siguiente paso evolutivo lógico de aplicación de los contenidos vistos en el capítulo previo, con el objetivo de establecer un camino progresivo de conceptos fundamentales, trabajos innovadores existentes, metodologías de aplicación general a partir de lo visto y finalmente la aplicación sobre un caso concreto y realista de lo aprendido.

Además de la propuesta del caso de estudio y de las premisas inicialmente contempladas, se detallan las vicisitudes que fueron surgiendo durante el avance del proyecto y que se debieron necesariamente afrontar, analizar e intentar resolver, para alcanzar el objetivo propuesto.

El objetivo de este capítulo es plasmar las dificultades que suceden en el área del ML y de qué forma, las aplicaciones y metodologías pertenecientes a la técnica de **TL** hacen su entrada en escena para solventar cada uno de esos obstáculos. De este modo, se apunta a brindarle al lector una visión más acabada del ciclo resolutivo realista consistente en la definición y abordaje de una determinada problemática mediante el desarrollo de modelos de ML, donde muchas veces surge la imposibilidad de cumplir ciertos requerimientos indispensables para el proyecto y finalmente, la oportunidad de superar dichas dificultades gracias al análisis del conjunto de metodologías de TL susceptibles de ser aplicadas, así como la comprensión de las ventajas y desventajas propias de su implantación.

De esta forma se intenta otorgarle significación y mayor grado de utilidad práctica a las secciones precedentes de este trabajo, consolidando de los conceptos expuestos en el marco teórico (desarrollado en el **capítulo 2**) y las diferentes subcategorías de aplicación estudiadas (detalladas en el **capítulo 3**), así como también la aplicación concreta del marco teórico resolutivo general estudiado (descrito en el **capítulo 4**).

5.2 - Presentación del modelo y las premisas

Como campo de resolución applicativa se postuló el perteneciente al **área del cuidado de la salud (HCA)**, más específicamente el área correspondiente al **diagnóstico por imágenes (DIA)**, que permite a los profesionales de la salud observar con detalle el interior de cualquier parte del cuerpo humano otorgándoles la posibilidad de detectar posibles patologías, aspirando a diagnósticos o planes de tratamiento más adecuados y precisos. Esta elección estuvo motivada en el análisis y consideración minuciosa de factores insoslayables, como el notable avance tecnológico (tanto de recursos hardware como software) que han experimentado los instrumentos de diagnóstico por imágenes en los últimos años, a su vez acompañado por la masividad de su producción y distribución, lo que ha transformado en algo muy usual su disponibilidad y manejo diario en diferentes centros de atención. Además se consideró el impacto beneficioso que su utilización genera sobre gran cantidad de personas así como también la creciente atención que ha recibido el HCA por parte de los gobiernos de la mayoría de los países, como resultado del surgimiento de la pandemia originada por el virus SARS-Cov-2 a fines del año 2019 y que lamentablemente persiste al momento de la redacción de este trabajo.

Considerando lo dicho previamente, se resolvió como la mejor opción concretar la implementación de un modelo clasificador tradicional dentro del paradigma de aprendizaje supervisado (**SLP**). Y como siguiente paso se procedió a definir una serie de objetivos o premisas para el desarrollo del modelo de ML propuesto, algunas de las cuales se enuncian a continuación:

1. **Precisión**, generar un modelo de clasificación robusto, que alcance un elevado rendimiento y precisión, considerando a priori y de manera referencial aproximada, una tasa de acierto base de clasificación aceptable superior al 90 %.
2. **Conjunto de datos**, relacionado al punto previo, disponer de un conjunto de datos compuesto de imágenes etiquetadas para las etapas de entrenamiento y validación de, al menos, 10.000 (diez mil) instancias. Considerando factible la utilización de la técnica de **augmentación de datos** para facilitar su obtención.
3. **Gratuidad**, implementación completamente gratuita; ya que el proyecto surge con fines investigativos y apunta a servir de guía para quien desee resolver cualquier problemática de ML imponiendo la menor cantidad de restricciones posibles. El conjunto de datos necesario, debía disponerse previamente desde algún repositorio de dominio público, sin costo alguno.
4. **Accesibilidad**, ligado al punto previo, el conjunto de datos de trabajo debe ser de acceso público y sin restricciones de privacidad que dificulten o imposibiliten su utilización y manipulación para fines académicos.

Cabe destacar que al momento de definir estas premisas, si bien el ámbito o área resolutive del problema a resolver estaba definida (y correspondería al campo de diagnóstico por imágenes dentro de **HCA**) y la técnica a utilizar era la de clasificación supervisada (**SLP**), como se mencionó previamente; no sucedía lo mismo con la tarea (objetivo) puntual a resolver, disponiendo de cierto margen de elección en esta instancia de avance del proyecto. En otras palabras, se pretendía determinar resolver cualquier problema del área HCA con la restricción de que cumpliera las premisas previamente definidas.

5.3 - Problemática afrontada

En el contexto del **ML**, se ha estudiado que el éxito de los **modelos** se encuentra directamente relacionado a la disponibilidad de un abundante conjunto de datos, así como al estrecho grado de dependencia respecto de estos para alcanzar un elevado rendimiento asociado al nivel de abstracción y generalidad de los modelos. Para determinadas tareas o dominios, dichos datos se encuentran previamente disponibles ya que han sido recopilados por empresas u organizaciones a lo largo de varios años de puesta en producción de sus aplicaciones; pero en ocasiones, los datos suelen ser propietarios y de difícil adquisición (en términos de tiempo, procesos de análisis, conocimientos requeridos, costo de insumos demandados, etc.). Quedando de manifiesto que los datos tienen un rol central en el esfuerzo de las organizaciones para su adquisición masiva y análisis minucioso a fin de transformarlos en información fundamental para la toma de decisiones que le permitan alcanzar el éxito, de allí la explosión de ciertas áreas relacionadas al ML como la minería de datos, analítica de grandes datos, análisis estadístico, visualización y ciencia de datos en general.

Después de una extensa búsqueda y el recorrido de una considerable cantidad de repositorios de imágenes emergió esta dificultad, que a priori resultaba cancelatoria del proyecto, dado que representaba la imposibilidad de cumplir concretamente con las premisas definidas (puntualmente las asociadas al tamaño y la gratuidad del conjunto de datos, descritas en la sección previa).

Los conjuntos de datos disponibles en el área escogida (diagnóstico por imágenes en HCA), representados como instancias de imágenes etiquetadas, contaban a lo sumo con algunos cientos de instancias, y si bien existían repositorios con mayor cantidad de instancias, los mismos eran privados y de acceso restringido a las organizaciones propietarias, mayormente nosocomios o clínicas privadas. Por consiguiente, no valía la pena abocarse a la construcción de un modelo de ML utilizando tan pocos datos porque esto amenazaba seriamente la probabilidad de obtener un modelo robusto de alto rendimiento, imposibilitando el cumplimiento de la premisa de precisión descrita en la sección previa y objetivo central del proyecto emprendido.

El paradigma tradicional de **aprendizaje supervisado (SLP)** inicialmente elegido para el desarrollo del proyecto, no se desempeña de forma satisfactoria en situaciones donde no se dispone de suficientes datos etiquetados para la tarea o dominio (objetivo) que se intenta resolver. Ya que es a partir de dicho conjunto de datos insuficiente, desde el cuál se debe generar en sucesivas etapas de entrenamiento y validación, un modelo de ML confiable, es decir, uno que minimice el error predictivo y maximice la tasa de aciertos (precisión).

5.4 - Solución propuesta

Ante la situación que se viene describiendo desde secciones previas, podemos evaluar que nos encontramos en una especie de “camino sin salida” ya que estamos ante la imposibilidad real de cumplir las premisas planteadas dentro del paradigma tradicional de ML definido (**SLP**) y en el área de implantación seleccionada (**HCA**). Ante esta encrucijada, las posibilidades aún disponibles para evitar la eventual cancelación del proyecto eran:

- Modificar las premisas planteadas, intentando flexibilizarlas o relajarlas de algún modo.
- Modificar el paradigma resolutivo de ML tradicional seleccionado (supervisado, SLP).

Ahora bien, modificar las premisas definidas inicialmente no resultaba una opción viable ya que uno de los objetivos del proyecto era la obtención de un clasificador de alto rendimiento que hiciera valer la pena su construcción y asociada a ella estaba la cantidad de datos que el modelo utilizaría. En este punto podemos notar que existen dos subgrupos entre las premisas, las asociadas a la **precisión** y el **conjunto de datos**, que revisten mayor importancia, considerándose inamovibles y por otro lado las asociadas a la **gratuidad** y la **accesibilidad** de los datos, que revisten menor importancia en comparación, por ende admitiendo ser flexibilizadas en algún aspecto.

Sumando a esto, ya hemos visto al estudiar los trabajos más sobresalientes del **capítulo de estado del arte**, que mucho del éxito y del factor novedoso de los mismos es justamente que no admiten premisas poco realistas, sino que por el contrario, plantean soluciones innovadoras a problemáticas concretas del mundo actual, eliminando restricciones ficticias. Siguiendo ese lineamiento, las premisas dispuestas por este proyecto resultaban lógicas para una aplicación de ML llevada a cabo por cualquier investigador o profesional de IT con recursos limitados, en resumen, no se trataba de premisas poco realistas y no se justificaba su manipulación. Finalmente, nada nos impedía evaluar otros paradigmas resolutivos distintos de los propuestos por el ML tradicional, que nos permitieran la utilización de diferentes conjuntos de datos para el entrenamiento, sin alterar el dominio de implementación ni las premisas inicialmente planteadas.

Es justamente ante este tipo de dificultades que el campo del TL emerge y evoluciona de manera sostenida día tras días a una velocidad asombrosa. Las diferentes técnicas y enfoques del área de transferencia del conocimiento nos permiten lidiar con escenarios como el propuesto, mediante el aprovechamiento inteligente de datos previamente etiquetados y disponibles gracias a la resolución de problemáticas sobre dominios relacionados al que se desea abordar (objetivo), así como también la utilización reflexiva del del conocimiento adquirido y almacenado, a fin de aplicarlo de manera ingeniosa para la asistencia del aprendizaje de nuestra tarea objetivo novedosa.

Por consiguiente, y gracias al conocimiento que el desarrollo del presente trabajo de investigación me ha otorgado, la idea fue analizar la aplicabilidad de diferentes metodologías actualmente existentes en el área de transferencia de conocimiento (**TL**), muchas de ellas estudiadas en el **capítulo de estado del arte**, con el objetivo de aprovechar sus beneficios para finalmente sortear las dificultades mencionadas, que nos permita abandonar esta situación de “camino sin salida” y alcanzar la solución buscada para la problemática definida.

De esta forma, recurriendo a técnicas de TL, observamos de qué manera ciertas restricciones, a priori postuladas como inamovibles para cualquier proyecto, pueden resultar de difícil cumplimiento práctico (como en este caso de estudio puntual donde resultaron directamente irrealizables) pero el modelo generado puede ahora relajarlas, siendo posible ya no depender de la disponibilidad u obtención costosa de un gran conjunto de datos etiquetados previamente, sino del aprovechamiento del conocimiento útil generado sobre otra tarea y dominio relacionado. Desde otra perspectiva, modificar el paradigma resolutivo nos permitirá sortear inteligentemente las dificultades de las premisas asociadas a la gratuidad y accesibilidad, que ya no serán necesarias sobre el dominio de aplicación seleccionado y cumplir con las restantes premisas fundamentales.

Se desplaza el foco de atención desde el conjunto de datos hacia el conocimiento que ya disponen otros modelos relacionados; logrando mediante técnicas de TL flexibilizar las imposiciones restrictivas inherentes al ML tradicional. Esta transferencia del conocimiento previamente recabado nos permite ampliar el espectro de aprendizaje utilizable hacia otras áreas donde los datos no son abundantes, gratuitos o de dominio público, posibilitando además que pequeñas empresas o investigadores que no disponen de estos enormes conjuntos de datos imprescindibles para el ML tradicional, implementen soluciones de alto rendimiento; lo cual resume el aspecto global, inclusivo y diversificador que el TL ofrece para todos aquellos investigadores, estudiantes u organizaciones pequeñas que deseen implantar soluciones a diferentes problemáticas del mundo actual.

Pasando al plano aplicativo y a fin de incrementar el grado de transferibilidad de las representaciones en el dominio origen (D_S) entre ambos dominios involucrados, deseáramos que las mismas fuesen lo más similares posibles para no tomar en cuenta características específicas que al ser discordantes respecto de los datos reales del dominio objetivo (D_T), pudieran perjudicar el proceso de transferencia y generen la aparición de **NT**. En este contexto, el enfoque destacado del TL, conocido como **adaptación de dominios (DA)** ayuda a salvar las discrepancias entre los datos de entrenamiento y pruebas a partir de métodos que generalmente buscan identificar características compartidas o aprender representaciones lo suficientemente generales como para ser aplicadas útilmente a ambos dominios. Uno de sus principales objetivos es la prevención de **NT** que se produce no sólo cuando los datos de entrenamiento resultan de escasa o nula utilidad para el aprendizaje sino tal vez contraproducentes, en el sentido que el modelo resultante entrenado a partir de datos con un grado de relación limitado con D_T puede exponer niveles de desempeño inferiores a los que evidenciaría si se hubiese entrenado prescindiendo de la transferencia de conocimiento.

Por lo expuesto previamente, sería indicada la aplicación de un método correspondiente al subárea de la **DA** en el cual, considerando cuidadosamente D_T y analizando su similitud relacional con los posibles dominios de origen desde los cuales efectuar la transferencia; pueda enfocarse en la selección de las instancias de datos que resulten más beneficiosas para el proceso de aprendizaje generativo del modelo objetivo. Por lo descrito en el capítulo del marco teórico (**capítulo 2 - Aprendizaje por transferencia**) y los trabajos analizados que representan el estado del arte de esta investigación (**capítulo 3 - Estado del arte**), sabemos que este procedimiento producirá un incremento significativo del rendimiento del modelo de ML final y posibilitará el alcance de la premisa fundamental de precisión postulada para este caso de estudio.

Podemos apreciar que el problema central a resolver aquí, está representado por un dominio aplicativo (diagnóstico por imágenes en **HCA**) donde es muy probable disponer de un conjunto de datos cuyas instancias etiquetadas resulten insuficientes para la generación de un modelo robusto de alto rendimiento desarrollado bajo el paradigma de ML tradicional (más precisamente **SLP**). Por consiguiente una de las alternativas viables dentro del TL es la aplicación del método de **“Aprendizaje por transferencia mediante reducción de dimensionalidad”** (analizado en la **sección 3.8**) que aplica la técnica de transferencia transductiva (**TTL**) mediante el enfoque de transferencia de representación de características, la cuál simplemente requiere disponer de un conjunto de datos de datos de entrenamiento que posea abundantes instancias etiquetadas e incluso admite que D_T no posea datos etiquetados, situación que si bien no acontece en nuestro caso de estudio (disponiendo de algunas instancias etiquetadas), posibilita que más información útil pueda extraerse, elevando la probabilidad de alcanzar un mayor rendimiento del modelo resultante. [51]

Se debe considerar que dentro del problema de clasificación supervisada planteado por este caso de estudio, las tareas involucradas deben estar relacionadas ya que el método a aplicar es un caso de **HetTL** donde únicamente se admite la transferencia de conocimiento entre modelos que posean diferentes dominios. En resumen, sobre la restricción de contar con un abundante conjunto de datos etiquetados para la etapa de entrenamiento, se añade la de tener el mismo espacio de etiquetas ($Y_S = Y_T$) y distribuciones de probabilidad condicional asociada ($P(Y_S | X_S) = P(Y_T | X_T)$) que en D_T . Se permite, por otro lado, que tanto los espacios de características ($X_S \neq X_T$) así como las distribuciones de probabilidad asociadas ($P(X_S) \neq P(X_T)$) en ambos dominios involucrados (D_S y D_T) en el proceso de transferencia de conocimiento difieran, aunque los mismos deben estar relacionados para posibilitar la extracción de un espacio latente de variables comunes, una de las tareas fundamentales que persigue la aplicación de este método. [27]

Si bien para el caso práctico que se intenta resolver, es factible la disponibilidad de un conjunto de datos previamente etiquetados pero demasiado escaso y por ende, insuficiente para la generación de un modelo de ML de alto rendimiento y la técnica de transferencia mediante reducción de dimensionalidad mencionada ha demostrado ser efectiva para guiar el aprendizaje y transferencia del conocimiento; debe destacarse la posibilidad de recurrir incluso a técnicas de aprendizaje autodidacta (**StL**) como la analizada en el trabajo **“Aprendizaje autodidacta: TL desde datos sin etiquetar”** (visto en la **sección 3.4**), que flexibilizará aún más el uso de diferentes conjuntos de datos, trayendo a consideración la oportunidad de entrenar el modelo utilizando datos no relacionados e incluso aquellos sin etiquetar, que abundan en internet actualmente.

De esta forma, podemos ir percibiendo de qué manera la aplicación de una amplia gama de técnicas dentro del TL permiten distender gradualmente las restricciones impuestas sobre los datos.

Resulta interesante destacar como la aplicación de ésta técnica de transferencia autodidacta citada previamente, nos permite prácticamente “mudarnos” entre paradigmas resolutivos (desde el supervisado hacia el no supervisado, **SLP** → **ULP**), lo cuál transforma drásticamente el escenario y nos permite migrar desde un ambiente hostil y desértico, en lo referente a la disponibilidad de datos etiquetados y costo asociado a su recopilación, para introducirnos en un ambiente fértil y mucho más versátil con posibilidad de sortear las dificultades previas para alcanzar las metas propuestas, disponiendo únicamente de datos sin etiquetar. Plasmando esta situación al caso de estudio concreto que nos atañe, ahora sería posible, por ejemplo, utilizar cualquier imagen radiográfica, tomográfica, de ultrasonido o resonancia magnética disponible en internet, sin importar que se relacione o no con el dominio de nuestra problemática puntual, limitándose únicamente al dominio de diagnóstico por imágenes HCA. Para clarificar con un ejemplo puntual, si nuestra tarea objetivo era la detección de tumores cerebrales mediante imágenes de resonancias magnéticas, ya no requerimos únicamente de éstas para entrenar y evaluar al modelo sino que ahora podemos incluso hacer uso de resonancias de cualquier otra parte del cuerpo, lo que incrementa ostensiblemente la cantidad de datos utilizables y que a su vez impactará en un mayor rendimiento del modelo.

Además, al poder prescindir de datos etiquetados, indirectamente se resuelve el problema asociado a la abundancia de instancias requeridas, dado que sabemos que en internet existen numerosas imágenes de este tipo, incluso siendo este dominio (el de diagnóstico por imágenes) muy específico. En síntesis, el paradigma de aprendizaje no supervisado revela un conjunto de posibilidades y herramientas que permiten ampliar el abanico de acciones resolutivas al relajar las rígidas restricciones que se presentaban sobre **SLP**, contemplando al mismo tiempo, la aplicación de soluciones que utilicen datos cada vez menos relacionados.

En este punto es oportuno aclarar que disponer de una mayor cantidad de instancias de datos etiquetadas sobre D_T tenderá a maximizar el rendimiento alcanzable por el clasificador construido. En ese sentido, si la cantidad de ejemplos etiquetados en D_T es muy escasa, existe la alternativa de incrementarla aún más mediante la aplicación de técnicas de **Data+**, muchas de las cuales han sido detalladas en el **anexo A**. Si bien la aplicación de estas técnicas trae aparejada una etapa extra de preprocesamiento de datos, se asume que su costo en tiempo y esfuerzo no será significativo, en base al reducido número de instancias que se pretende aumentar y contrastándolo con el impacto positivo que este proceso puede generar sobre el modelo de ML final. Como ejemplo ilustrativo podemos considerar la forma en que el algoritmo de **codificación dispersa (SCA, visto en la sección 3.4)** para el aprendizaje autodidacta o el **algoritmo de aprendizaje selectivo (SLA, visto en la sección 3.11)** permiten contemplar conjuntos de datos abundantes ya sea recurriendo a distintos paradigmas o al aprovechamiento de dominios escasamente relacionados.

Por su parte, el grado de similitud que exista entre dominios y tareas involucrados en el proceso de transferencia será determinante, es decir, es muy importante que las instancias de datos de entrenamiento sin etiquetar (abundantes) disponibles en D_S y las de pruebas etiquetadas (escasas) disponibles en D_T conserven un alto grado relacional. De manera análoga y complementaria a lo expresado en el párrafo previo, resulta evidente que cuanto más se asemejen las instancias de los dominios intervinientes en la transferencia, mayor será su nivel de representatividad, permitiendo al modelo objetivo extraer información más útil desde ellas en etapa de entrenamiento y por consiguiente, alcanzar mayor precisión y eficiencia. Como ejemplo ilustrativo podemos considerar la forma en que el algoritmo **TrAdaBoost** (visto en la **sección 3.3**) aprovecha un reducido conjunto de instancias etiquetadas para maximizar el rendimiento o del método de **Rtm** (visto en la **sección 3.10**) que bajo el paradigma **RTL** puede aprovechar el grado de relación entre tareas involucradas para efectuar una transferencia más eficiente, que decantará en la generación de modelos de mayor calidad.

El amplio abanico de metodologías que admite el TL nos posibilita aplicar soluciones aún más extremas, considerando incluso como factible la aplicación práctica de la técnica propuesta por el trabajo “**TL sobre dominios distantes**” (analizada en la **sección 3.11**), que permitiría ampliar el espectro de posibles dominios de entrenamiento correspondiente al conjunto de datos previamente etiquetados. Permittiéndonos alejarnos del relativamente restrictivo dominio aplicativo del cuidado de la salud (**HCA**) sorteando las dificultades de disponibilidad (adquisición, privacidad, libre acceso y gestión) de datos asociadas al mismo, para migrar desde un escenario de dificultad relativamente alta hacia otros novedosos donde vicisitudes de este tipo ya no se presenten. La posibilidad de aplicar estrategias de este tipo demuestran que incluso dentro del paradigma **SLP**, es posible intentar obtener un conjunto de datos previamente etiquetados y con gran cantidad de instancias pero ya no necesariamente dentro del dominio HCA, sino de cualquier otro menos restrictivo que sabemos que existen actualmente en internet, siendo muy útiles y variados.

Profundizando lo mencionado en el párrafo previo, aplicando esta técnica podemos hacer uso de conjuntos de datos masivos de acceso público que pertenezcan a dominios no relacionados. Para ejemplificar, podemos citar algunos entre los que se encuentran: **EMINIST**, una extensión del conocido **MINIST** publicada en 2017, que contiene alrededor de 280.000 imágenes de dígitos y caracteres escritos a mano. Naturalmente y como se mencionó, cuanto mayor cantidad de instancias posea el conjunto de datos que se utilizará para el entrenamiento y validación, mayor será la probabilidad de alcanzar un modelo objetivo robusto de alto rendimiento y el cumplimiento de una de nuestras premisas iniciales. Otros conjuntos de datos muy utilizados son **ImageNet**, organizado en nodos donde cada uno contiene miles de imágenes de libre acceso para investigación y **COCO**, un conjunto de datos de gran escala (con más de 200.000 imágenes etiquetadas) para la comprensión de imágenes genéricas. [25] [50]

Otra alternativa resolutive que es apropiado mencionar, es la que posibilitan las herramientas de desarrollo actuales para el campo del ML (como Keras, PyTorch, TensorFlow o Theano, entre otras) donde es posible hacer uso de ciertos modelos previamente entrenados y posteriormente aplicarlos sobre un determinado dominio objetivo relacionado. Se trata, al fin y al cabo, de la implantación similar a las alternativas vistas previamente pero bajo un nivel de abstracción superior haciendo uso de las herramientas disponibles para la generación, desarrollo y refinamiento de modelos de ML. La idea es echar mano de la utilización de un modelo de ML ya creado y entrenado con abundantes instancias de datos y reutilizarlo sobre el dominio que se pretende resolver. De hecho, existen trabajos de tesis de grado presentadas en la facultad de informática (**UNLP**) donde se aplican soluciones mediante la implantación de esta metodología.

Por lo visto a lo largo del desarrollo del presente capítulo, la implementación de las técnicas de TL esgrimen entre sus principales virtudes, la posibilidad de relajar las restricciones de disponibilidad propias de los conjuntos de datos sobre el dominio objetivo, requerimiento primordial y muchas veces difícil de cumplir como en el caso de estudio presentado. Para posteriormente abocarse a solventar esta situación mediante el aprovechamiento del conocimiento previo disponible y generado sobre dominios de origen relacionados. De esta forma es factible la generación de un clasificador tradicional supervisado disponiendo de un conjunto de datos acotado en cuanto a cantidad de instancias, evitando un incremento en los costos constructivos del modelo y conservando al mismo tiempo, un elevado rendimiento de los mismos.

5.5 - Conclusión

En este capítulo se detalló brevemente la idea de resolver un problema de ML aplicado a un caso práctico del mundo actual, propio del ámbito del cuidado de la salud relativo al diagnóstico de patologías por imágenes. Describiéndose el recorrido de las etapas de desarrollo del proyecto donde se suscitaron dificultades que impedían el cumplimiento de las premisas establecidas, amenazando la concreción del mismo.

El análisis de un caso de estudio realista, reviste gran utilidad y relevancia ya que permite exponer las dificultades resolutivas de una problemática mediante el modelado de ML y clarifica no sólo de qué manera el TL puede utilizarse para superar estas dificultades, sino además de qué forma la aplicación de diferentes metodologías posibilitan relajar restricciones previas y abren un abanico de alternativas para su resolución.

En síntesis, el abordaje del capítulo actual representó la posibilidad de aplicación concreta de uno o más métodos vistos en el marco teórico para un problema determinado, y resultó fundamental para que el lector obtenga una perspectiva más abarcativa e integradora asociada al área del TL, donde se complemente el contenido fundamental visto en el plano teórico con una descripción de los pasos resolutivos sobre un caso de estudio en el plano práctico.

Capítulo 6 - Aportes, conclusiones y trabajos futuros

6.1 - Introducción

En este capítulo, se describirán los aportes que el desarrollo de esta tesina de investigación ha podido concretar, así también como las conclusiones a las que se ha arribado como resultado del proceso de análisis efectuado. Finalmente, se establecerán los lineamientos de trabajo a futuro que han aflorado al enfrentar las dificultades propias de cada metodología propuesta o bien que representan mejoras u optimizaciones propuestas cuya implantación representaría un valioso avance para el campo del TL.

6.2 - Aportes

Esta tesina proporciona los contenidos necesarios y fundamentales para quienes estén dando sus primeros pasos en el área de ML, más puntualmente en el TL; gracias al análisis de su surgimiento, exposición de conceptos y definiciones fundamentales, tópicos que pretende resolver y problemáticas propias de su aplicación, esquemas metodológicos, el estudio de un caso aplicativo y la implantación de todo lo aprendido, además de distintos aspectos complementarios abordados en secciones anexas, un extenso glosario de definiciones con referencias a fuentes informativas y referencias externas a la bibliografía utilizada a lo largo de esta investigación. Por lo previamente mencionado, el presente trabajo se plantea como el peldaño inicial de un camino apasionante, ramificado en múltiples subcampos atractivos y que representa un crecimiento invaluable para cualquier profesional de la informática o interesado en estos temas.

Si bien la presente tesina se ubica dentro del área de investigación teórica, pretende aportar no sólo una serie de conceptos y definiciones básicos respecto del TL, sino además consolidar un recorrido pragmático focalizando en el lector que comienza a incursionar en la temática, con el objetivo de ampliar sus herramientas de cara a la investigación, el emprendimiento y desarrollo soluciones que puedan obtener el máximo desempeño sacando provecho de esta técnica y evitando al mismo tiempo errores frecuentes e incluso situaciones indeseadas (como **NT** o **CoD**).

A partir del conocimiento de los principales trabajos abordados y el análisis de un caso de estudio realista al que nos hemos enfrentado, hemos podido establecer de qué forma el TL puede solventar dificultades concretas a las que se enfrenta el ML y que pueden transformarse en limitantes de proyectos innovadores. De este modo, el lector añadirá al conocimiento teórico ofrecido, experiencias y situaciones realistas a fin de dimensionar las dificultades de cada escenario, trazar paralelismos con la problemática concreta a la que se enfrenta y encaminarse hacia las metodologías resolutivas de TL más adecuadas.

Por último, este trabajo ha puesto de manifiesto diversos factores de mejora u optimización sobre los métodos investigados, donde la idea es que el lector pueda reflexionar y avanzar sobre cualquiera de ellos o incluso la proposición de ideas más radicales que desencadenen en métodos superadores a los descritos aquí, suplantando a los que actualmente representan el estado del arte de cada subárea puntual.

6.3 - Conclusiones

El desarrollo de la presente tesina ha permitido realizar una introducción y profundización de los principales tópicos relacionados al área del **TL** dentro del campo del **ML**. De esta forma se ha procedido a la examinación de diferentes factores determinantes a tener en cuenta, distintos contextos realistas y motivaciones que han impulsado su surgimiento y consolidado su avance.

Se han proporcionado conceptos, definiciones formales, categorización en técnicas y enfoques más sobresalientes del área, acercándole al lector de este documento, la posibilidad de análisis, consideración y comprensión de los factores más determinantes propias de las metodologías de resolución adecuadas y la evaluación de las diferentes vicisitudes que acarrea cada contexto y problemática afrontada que facilite el discernimiento de las metodologías más indicadas así como limitaciones o eventuales riesgos aplicativos de las técnicas de TL.

El abordaje de tópicos novedosos tratados aquí, permitirán al lector incrementar considerablemente el número de herramientas fundamentales que puede esgrimir ante cada problema que desee afrontar y resolver haciendo uso del TL. A tal fin se han expuesto algunas de las más innovadoras y relevantes metodologías e investigaciones presentes en el área, que posibilitan el manejo de un panorama más abarcativo de las distintas estrategias resolutivas disponibles y su aplicabilidad ante cada situación del mundo actual.

Se ha planteado un caso de estudio realista, basado en mi experiencia reciente incursionando en el campo del ML, con el objetivo de permitirle a los lectores mensurar el valor que representa el conocimiento de las metodologías resolutivas inherentes al TL ante las barreras que pueden surgir dentro del campo del ML y que necesariamente deben ser progresivamente superadas por parte de investigadores y desarrolladores. También se ha propuesto un marco metodológico que propone una secuencia resolutiva para afrontar problemas del mundo real mediante TL, y que representa la integración de los contenidos expuestos.

Finalmente, a partir del recorrido realizado a través de los trabajos más relevantes del estado del arte en lo referente al TL, fue posible exponer abundantes líneas de investigación futuras que evidencian el grado de proyección de cada subárea, justificando su progreso acelerado y sostenido en el tiempo.

6.4 - Trabajos futuros

Los seres humanos poseen mecanismos de decisión que determinan cuándo transferir información, identificando fuentes de conocimiento y nivel de abstracción adecuados, el problema de determinar la manera de tomar esas decisiones en un contexto generalizado de ML, particularmente evitando problemáticas entre las que se encuentra la **NT** y proponiendo metodologías con un nivel de automatización elevado, por ejemplo al establecer mapeos entre diferentes tareas, confieren desafíos de relevancia a resolver en el futuro. [2]

Otro desafío a futuro consiste en permitir la transferencia entre tareas que representen un amplio grado de diversidad, es decir, donde exista un muy bajo grado relacional mutuo. En este aspecto, el trabajo titulado **“Transferencia profunda mediante lógica de Markov de segundo orden”**, apunta al descubrimiento de patrones sobre una tarea de origen para hallar fórmulas de segundo orden que representan conceptos abstractos universales como simetría y transitividad. Este trabajo pertenece al enfoque **ITL**, siendo el conocimiento de la tarea origen muy abstracto y admitiendo que ambas tareas difieran significativamente. [30]

Otro reto a resolver consiste en realizar la transferencia sobre conjuntos de pruebas que revistan mayor complejidad. Particularmente en el área del **RTL**, donde la transferencia de conocimiento puede tornarse mucho más dificultosa a medida que las tareas involucradas se dificultan. Ante la probabilidad de afrontar aplicaciones de **RL** de alta complejidad, es importante no limitar la investigación a dominios simples. [2]

En el futuro será necesario avanzar sobre varios problemas de investigación cruciales y de resolución actualmente abierta dentro del campo del TL; tal es el caso de la interrogante: ¿cómo evitar la **transferencia negativa**? y el análisis de los factores intervinientes a considerar, como el grado de relación existente entre los dominios participantes en la transferencia o la capacidad de cada algoritmo para evitar su ocurrencia y de qué manera impactan dichas salvaguardas en su potencial del rendimiento. [1]

Muchos de los algoritmos de transferencia existentes se focalizan en el incremento del grado de generalización a través de dominios disímiles, donde espacios de características o distribuciones de probabilidad asociadas pueden diferir considerablemente, entonces los métodos de transferencia aprovechan el conocimiento compartido entre dichos dominios así como el descubrimiento de relaciones mutuas para la construcción del sesgo inductivo del dominio de la tarea. En este aspecto, el trabajo titulado “**Aprendizaje por transferencia heterogéneo para el agrupamiento de imágenes mediante la web social**” expone la utilización de técnicas mencionadas, refiriéndonos a este tipo de TL como **transferencia heterogénea (HetTL)**, que representa una línea de investigación futura relevante dentro del campo del **LML**. [3] [31]

Por su parte, la investigación relacionada al **LML** considera dos líneas de avance claramente marcadas, primeramente dentro de la comunidad **IA**, donde se lo ha considerado como un gran desafío que elevará el perfil de investigación del área con un amplio abanico aplicativo; en segundo lugar establecer un proyecto de código abierto que permita a investigadores compartir conocimiento y colaborar en sistemas de este tipo. [3]

También debemos considerar que el TL se ha convertido en uno de los subcampos protagonistas dentro del ML, esgrimiendo beneficios ideológicos al ser visto como un aspecto importante del aprendizaje humano y también como trampolín hacia enormes beneficios prácticos, sobre los cuales abundan las demostraciones empíricas (muchas de las cuales hemos recorrido en este trabajo) que han demostrado su potencial para aumentar de forma significativa la eficiencia de los algoritmos de ML.

A medida que aumenta la potencia informática y los investigadores aplican metodologías resolutivas del campo del ML sobre problemas de creciente complejidad, el TL puede sólo tender a ser más valorado y sus resultados muy deseables. Las técnicas más avanzadas han sido mayormente utilizadas sobre aplicaciones de pequeña escala y variabilidad limitada, tal como la localización basada en redes de sensores, la clasificación de textos e imágenes. Pero en el corto plazo, estas técnicas deberán ser ampliamente utilizadas para resolver aplicaciones más desafiantes, como la clasificación de video, análisis de redes sociales e inferencia lógica. [1] [2]

La serie de trabajos relacionados al TL analizados en el **capítulo de estado del arte**, representan no sólo algunos de los tópicos más novedosos y relevantes, sino que además plantean diferentes puntos a tratar en el avance de cada metodología, ya se trate de extensiones u optimizaciones, profundización de problemas poco conocidos o abordados y el análisis de situaciones donde los resultados empíricos aparentan ser contradictorios respecto de ciertos estudios teóricos que ameritan la profundización de investigaciones. A partir de estas situaciones, a continuación se mencionan brevemente cada una de estas posibles ramas de investigación futura, que como resultado de la lectura de esta tesina, resultará oportuno considerar.

En el trabajo titulado **“TrAdaBoost: un impulso para TL”**, se analizó que la mejora obtenida en la convergencia del error de predicción depende en gran medida de la calidad de las instancias de diferente distribución, este problema requiere mayor profundidad de análisis y comprensión; por otro lado vimos que la tasa de convergencia del algoritmo de transferencia **TrAdaBoost** puede ser comparativamente lenta y que la transferencia del conocimiento puede provenir desde una única distribución, no admitiendo el tratamiento simultáneo de múltiples distribuciones; de este modo, como línea de investigación futura se contempla su extensión a fin de superar las limitaciones mencionadas y obtener un entorno más robusto y abarcativo.

El trabajo titulado **“Aprendizaje autodidacta: TL desde datos sin etiquetar”**, ha planteado una interrogante teórica destacada y aún abierta: ¿de qué forma la “similitud” entre datos etiquetados y sin etiquetar afecta el rendimiento del **StL**? Por otro lado, se analizó cómo otros investigadores han propuesto métodos basados en la selección manual de heurísticas para generar problemas de predicción secundarios a partir de los datos sin etiquetar. A partir de esto podemos afirmar que la adaptación de esos métodos por parte de las aplicaciones de **StL**, representan un flujo de investigación aún con mucho trabajo por delante.

El trabajo titulado **“TL mediante mapeo estructural local de modelos múltiples”**, plantea la necesidad de continuar explorando diferentes estrategias de análisis situacional donde deberían combinarse diferentes dominios de origen en lugar de efectuar la construcción de modelos individuales. Además sería de utilidad efectuar comparaciones entre el modelo propuesto (**LWE**) y otros algoritmos de TL basados en un único modelo, así como desarrollar otros métodos efectivos que permitan la configuración de parámetros.

El trabajo titulado **“Mapeo y revisión de redes lógicas de Markov para TL”** por su parte, propone como línea de mejora, avanzar en la incorporación del marco de trabajo propuesto (**TAMaR**) sobre escenarios de transferencia novedosos mediante el desarrollo de técnicas de búsqueda de mapeos de predicados en dominios donde el número de predicados existente torna prohibitiva la implementación de una búsqueda exhaustiva, así como analizar y definir nuevas métricas de similitud entre dominios para determinar a priori y con relativa certidumbre, que tan beneficiosa puede resultar la aplicación de técnicas de TL.

Por su parte, en el trabajo titulado **“Aprendizaje y evaluación de clasificadores bajo el sesgo de selección de la muestra”**, hemos visto que, los árboles de decisión se desempeñan bien ante el **SSB** a pesar de ser catalogados como aprendices globales, obtenido resultados contrarios a los teóricamente esperados. Este hecho plantea la necesidad de avance en el estudio de los efectos del **SSB** sobre estos modelos así como los factores estructurales que pudieran explicar el comportamiento observado, para una mayor comprensión.

El trabajo titulado **“Aprendizaje por transferencia mediante reducción de dimensionalidad”**, ha evidenciado la necesidad de extender el algoritmo de transferencia propuesto (**MMDE**) mediante la extracción de características no negativas de modo que pueda ser de ayuda en la transferencia con otros clasificadores tradicionales; por otro lado se ha planteado la búsqueda de métodos novedosos de mayor eficiencia que extiendan a **MMDE** y posibiliten el abordaje de problemas de TL a gran escala.

Por su parte, el trabajo titulado **“Reducción de dimensionalidad transferida”** ha propuesto como línea de investigación futura, la optimización de la etapa de agrupamiento de la cual se compone al algoritmo propuesto (**TDA**) ante situaciones donde ni siquiera el número de clases de datos es conocida. Así mismo es deseable la incorporación de una mayor cantidad de tipos de conocimientos desde más dominios de origen, que permitan exponer con mayor claridad la relación entre diferentes conceptos.

El trabajo titulado “**Transferencia entre dominios para el aprendizaje por refuerzo**”, ha puesto de manifiesto que el análisis y estudio de las relaciones entre dominios a transferir dentro del RL representa un problema de relevancia y relativamente poco explorado, dicho trabajo representó un paso inicial en el aprendizaje del mapeo, ilustrando un proceso por el cual puede derivarse un conocimiento cualitativo de alto nivel. La mayoría de los trabajos previos se centraron en la codificación manual de funciones de mapeo entre dominios pero el aprendizaje de esos mapeos confiere un problema de investigación muy desafiante; a tal punto que actualmente la investigación gira en torno al desarrollo de un método de mapeo automático entre dominios y la construcción automatizada de tareas de origen que se beneficien del conocimiento transferido.

El trabajo titulado “**Transferencia de aprendizaje entre dominios distantes**”, plantea la necesidad de contemplar situaciones comunes en el mundo real y pocas veces abordadas como las que se relacionan con la transferencia de conocimiento cuando los dominios involucrados presentan un grado de relación muy lejano y como paso relevante de investigación futuro, se plantea la extensión del algoritmo propuesto (**SLA**) para incorporar la posibilidad de admitir el tratamiento de múltiples dominios de origen.

El desarrollo de esta investigación aborda diferentes situaciones y problemáticas relevantes que presentan “zonas grises” donde es evidente que ameritan más tiempo de investigación, estas áreas parcialmente exploradas no sólo existen dentro de las diferentes metodologías de TL sino también en sub áreas relacionadas (como **NT** y **LML**). Por consiguiente, un abordaje de cualquiera de estos tópicos reviste gran valor en el contexto del **ML**. Al mismo tiempo, cada trabajo presentado en el **capítulo de estado del arte**, además de focalizar sobre una determinada investigación abocada en resolver una problemática concreta, describe factores de mejora u optimización asociados, campos poco explorados, problemáticas inherentes, que pueden servir como paso inicial para un recorrido investigativo que impulse su desarrollo.

6.4.1 - Líneas de investigación propuestas

Como resultado del proceso de desarrollo y construcción de este trabajo, han emergido distintas líneas de continuidad investigativa y avance, entre las cuales podemos citar las siguientes:

- Extender el proceso metodológico desarrollado en el **capítulo 4**, ya sea abordando nuevas etapas no contempladas, extendiendo los procesos descriptos, incorporando alguna actividad novedosa o no contemplada, así como el refinamiento de cualquiera de los flujos presentados.
- Concretar la implementación del caso aplicativo descripto en el **capítulo 5**, utilizando alguna herramienta de desarrollo actual (como Keras, PyTorch, TensorFlow o Theano) o haciendo uso de modelos pre-entrenados, lo cual permitiría completar todo lo expuesto con resultados empíricos a fin de contar con una comprobación fehaciente del análisis teórico efectuado.
- Proponer un nuevo método de TL bajo alguna de las categorías y enfoques vistos que resulte de interés y solvente alguna de las problemáticas que pudieran presentar las metodologías del área. Si bien este punto requeriría conocimientos profundos y sólidos de matemáticas y estadística, su alta complejidad no debería tornar dicha tarea en algo imposible.
- Consolidar alguna de las mejoras propuestas en este capítulo (sección de **trabajos futuros**), representando cualquier avance sobre alguna de las “zonas grises” descritas, representaría un invaluable avance; como el análisis de algún punto débil en determinada metodología para proponer optimizaciones que signifiquen la evolución del algoritmo base en cuestión.

- Establecer un estudio comparativo entre dos o más métodos de TL donde se ponga de manifiesto ante qué contextos de aplicación cada técnica resulta más indicada o sobre cuales no resultaría apropiada, exponiendo justificaciones para cada caso a fin de clarificar de manera pragmática qué herramienta se debería aplicar o evitar, en base a un conjunto de problemáticas más usuales y realistas de la actualidad.
- Plantear la fusión de dos o más técnicas de TL, de modo de integrar los puntos fuertes que propone cada una y aprovecharlos para la optimización de un novedoso método combinado, describiendo que factores negativos de cada método se superan con la fusión, estudiando los contextos de aplicación en los que se maximizarían los beneficios de su utilización, así como los casos existentes que hagan uso de una combinación de diferentes técnicas dentro del área.
- Presentar un marco de trabajo innovador que ataque una variedad de problemáticas mediante la combinación de diferentes secciones resolutivas y generando un entorno de alto poder (notar que se trata de la combinación de proposición de un nuevo método y la combinación de ellos). A modo de ejemplo, podríamos considerar un entorno que permita resolver problemas asociados al **HetTL** abocándose a igualar inicialmente el espacio de características entre dominios para posteriormente atacar el nuevo problema de tipo **HomTL**, focalizando en la diferencia de distribuciones asociadas mediante la utilización de otras técnicas en ese subcampo.

6.5 - Conclusión

En este capítulo, se han expuesto los aportes que el desarrollo del presente trabajo ha podido plasmar, además de las conclusiones generales extraídas a partir de su construcción; por último, se ha detallado un conjunto relativamente amplio y variado de posibles senderos de investigación a futuro, que apuntan a construir soluciones innovadoras sobre áreas identificadas como poco exploradas o sencillamente a la optimización y refinamiento de las implementaciones actuales que conforman el estado del arte, apuntando a la profundización investigativa y el desarrollo continuado del área.

Las secciones del presente capítulo han contribuido a la integración de varios capítulos previos de esta tesina, especialmente el **capítulo de estado del arte**, a partir del cuál se han identificado y extraído áreas de investigación y desarrollo futuras, así como también del **capítulo de propuesta metodológica para la aplicación de TL**, proponiendo alternativas posibles para su extensión y refinamiento, y finalmente del **capítulo de estudio de un caso aplicativo**, para su implantación concreta utilizando las técnicas de TL propuestas; además de la proposición de otras ideas novedosas para futuros desarrollos y la enumeración de los aportes que éste trabajo humildemente ha realizado.

El abordaje y redacción de los temas expuestos en el presente capítulo resultó trascendental ya que representó una síntesis global abocada a brindarle al lector una visión panorámica del trabajo completo, permitiéndole comprender y dimensionar la relevancia que ha cobrado recientemente el área del TL, confiando, a partir del análisis de su acelerada evolución, en que continuará incrementándose de manera notable en el corto y mediano plazo.

Anexos

A - Aumentación de datos [21]

A.1 - Origen y definición del problema

A medida que los **modelos de ML** se tornan más complejos como resultado de su aplicación sobre una gama de problemas de dificultad incremental, la cantidad de datos requerida para la obtención de un rendimiento elevado, también se incrementa considerablemente. No obstante, si bien existen muchas áreas donde el caudal de información aumenta aceleradamente, no todos los dominios han visto este aumento en la disponibilidad de datos, en ocasiones debido al alto costo que significa la obtención de muestras etiquetadas.

Usualmente, los datos necesarios para el entrenamiento de los modelos de ML se presentan de forma desestructurada y conteniendo ruido (información no relacionada, poco relevante o redundante), en consecuencia, una de las tareas esenciales del proceso generativo consiste en la recolección y preparación de los datos a un formato adecuado para ser proporcionado como entrada de un algoritmo. Particularmente en el ámbito asociado al **aprendizaje profundo supervisado (SDL)** y **redes neuronales convolucionales (CNNs)**; el logro de elevadas métricas de rendimiento y precisión se asocia directamente a la disponibilidad de conjuntos de datos de gran tamaño y diversidad, utilizables durante la etapa de entrenamiento. Entonces, la problemática que potencia el surgimiento de esta técnica radica justamente en los elevados costos y dificultad de obtención de datos de calidad en cantidades significativas, que se ajusten a la tarea objetivo a resolver. [32]

La **aumentación de datos (Data+)** es una técnica de procesamiento de datos que combate la problemática asociada a su carestía, permitiendo que el conjunto original disponible sea ampliado a fin de que los modelos de ML construidos experimenten un incremento del grado de **generalización** alcanzable, mayor control del **sobreajuste** y finalmente niveles superadores de confiabilidad y rendimiento. En concreto, consiste en la generación artificial de nuevas instancias de datos por medio de perturbaciones realizadas sobre los datos originales, para abordar los requisitos de diversidad y abundancia de datos de entrenamiento así como también ayudar a balancear eventuales desequilibrios de datos entre las diferentes clases existentes.

La correcta aplicación de **Data+** posibilita el incremento cuantitativo y cualitativo del conjunto de datos a fin posibilitar la construcción de **modelos de ML** que minimicen la probabilidad de **sobreajuste**, haciendo que el error del modelo en la etapa de prueba (utilizando datos desconocidos) y en la etapa de entrenamiento (utilizando datos conocidos), converjan a medida que se incrementan las épocas o rondas de entrenamiento, marcando un elevado nivel de **generalización** del **modelo**, como lo ilustra a continuación la **figura A.1**.

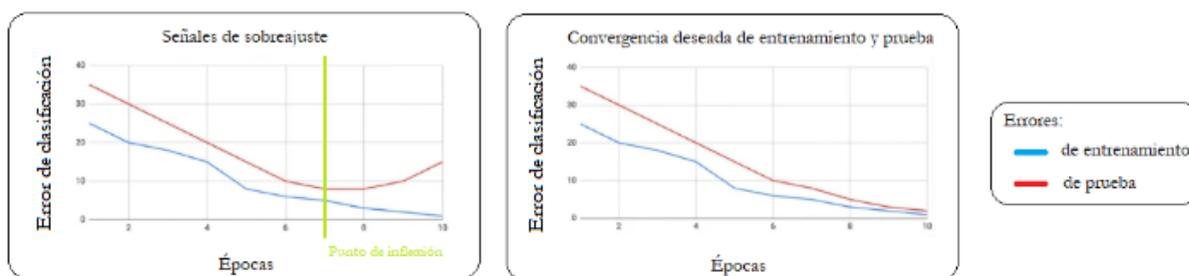


Figura A.1: Tasas de error de entrenamiento y prueba (sobreajuste vs. generalización)
Figura adaptada de **A survey on Image Data Augmentation for Deep Learning - Shorten y Khoshgoftaar**

A.2 - Técnicas

Como se mencionó previamente, en modelos de ML, la robustez del proceso de entrenamiento así como la precisión alcanzada dependen altamente de la disponibilidad de grandes volúmenes de datos etiquetados, los cuales son costosos de producir. Un enfoque eficaz para tratar con este problema es la utilización de métodos de **Data+** que posibiliten la generación automática de nuevas instancias etiquetadas.

Entre los dominios de mayor estudio y utilización de técnicas de Data+ podemos citar:

Aumentación de imágenes para aplicaciones de visión por computadora, estrategias para aumentar los datos de entrenamiento mediante transformaciones geométricas simples sobre las imágenes de entrada (volteos en diferentes sentidos, recortes, rotaciones, cambios de perspectiva, traslaciones, acercamientos o alejamientos, inyección de ruido, etc.), transformaciones sobre el espacio de color (fundido, brillo, opacado, contraste, saturación, aplicación de filtros, mezclado, borrado aleatorio, difuminado o ruido, etc.). Existen multitud de bibliotecas para tales fines como: PyTorch, ImGaug, Albumentations, entre otras.

Aumentación basada en entrenamiento adversario, donde el objetivo es transformar imágenes para “engañar” a un modelo de aprendizaje profundo (**DLM**) para compensar la debilidad del modelo original. El método aprende a generar máscaras las cuales son aplicadas a imágenes de entrada a fin de permitir la generación de nuevas imágenes aumentadas.

Aumentación basada en redes generativas adversarias (GANs), consta de dos redes neuronales artificiales (**ANNs**) entrenadas simultáneamente, compuestas por un generador cuyo objetivo es generar imágenes falsas (construidas artificialmente) a partir del espacio latente y un discriminador cuyo objetivo es distinguir las imágenes falsas sintéticas generadas, de aquellas imágenes reales.

Aumentación basada en transferencia de estilo neuronal, mediante **ANNs** se genera una imagen aumentada a partir del contenido extraído desde una determinada y el estilo desde otra diferente; entonces la imagen generada es combinada para asemejarse a la imagen de entrada pero pintada con otro estilo diferente.

Técnicas de aumentación textual para aplicaciones NLP, la minería de textos, a diferencia de la de datos, trata con datos no estructurados; por esa razón los métodos de esta área suelen incorporar una fase previa de análisis y “estructuración” del contenido para permitir la posterior aplicación de algoritmos clásicos de minería de datos. Además se debe tener especial cuidado ya que modificaciones inadecuadas podrían corromper los datos alterando su significado. Suelen utilizarse heurísticas de aumentación, como traducciones, reemplazo por sinónimos, eliminación de adjetivos, alteración de oraciones y fragmentación de revisiones. [46]

B - Reducción de dimensionalidad [17] [18] [48] [49]

B.1 - Origen y definición del problema

El **ML** se utiliza para la resolución de tareas que suelen poseer un conjunto de datos sobrecargado en un espacio de alta dimensionalidad (en el orden de cientos o miles de atributos), incrementando costos, complejidad y tiempo del proceso de construcción y entrenamiento de los **modelos**, además de favorecer el **sobreajuste** y deterioro del rendimiento.

Este problema trae aparejadas consecuencias similares a las de la carencia de datos (vista en el **anexo previo**) en el siguiente aspecto, asumiendo una representación tabular del conjunto de datos (donde cada fila representa una instancia y cada columna un atributo del espacio de características), mientras que aquel problema visto se asociaba a la escasez de filas, éste se asocia a la sobreabundancia de columnas.

En el contexto de ML y estadísticas, la reducción de dimensionalidad, designa a la etapa de preprocesamiento de datos (previa a la construcción y entrenamiento de modelos) que consta del análisis y manipulación de los mismos a fin de reducir el número de variables aleatorias, prescindiendo de todos aquellos datos irrelevantes y redundantes para la tarea (objetivo) a resolver. Por consiguiente, se produce el incremento de la precisión del aprendizaje, la interpretabilidad de los resultados y la simplicidad tanto de los datos como del modelo generado. Este proceso consiste en la asignación del espacio de características original sobre uno de dimensionalidad restringida mediante la selección de un subconjunto de ellas o la generación de otro que posea mayor expresividad.

Entre las principales motivaciones que alientan la reducción de dimensionalidad, consideramos:

1. Reducción del conjunto de datos de formación original, conservando su relevancia, variabilidad y expresividad informativa.
2. Reducción de los requerimientos de información iniciales, minimizando costos formativos.
3. Incremento del poder predictivo (velocidad y precisión) de los modelos de ML generados.
4. Aumento de la eficiencia y velocidad de ejecución del proceso de aprendizaje, potenciado por la aceleración de los algoritmos de ML utilizados.
5. Mayor comprensión de los datos y simplificación de los resultados, facilitando el entendimiento de los procesos subyacentes mediante metodologías de visualización de datos.
6. Compactación y refinamiento de la representación de los datos, simplificando su manipulación.
7. Eliminación de atributos engañosos, redundantes (correlación) o irrelevantes (ruido) para la tarea abordada; incrementando la calidad de los datos que simplifica y optimiza los modelos.
8. Ampliación del espectro de algoritmos de entrenamiento aplicables, no contemplados previamente debido al bajo desempeño que presentan sobre datos de alta dimensionalidad.
9. Eliminación de la multicolinealidad y sobrecarga de algoritmos, que decanta en la simplificación e incremento de la capacidad de generalización de los modelos, reduciendo la probabilidad de sobreajuste y optimizando su rendimiento sobre datos realistas desconocidos (datos de prueba).

Entre las desventajas que puede acarrear la fase de implementación de reducción de dimensionalidad, podemos mencionar que la aplicación de ciertas técnicas pueden ocasionar la disminución o incluso la pérdida de interpretabilidad de los datos, sumado a que el proceso en sí mismo representa una etapa extra de preprocesamiento de datos que añade costos y cierto esfuerzo que debe ser considerado.

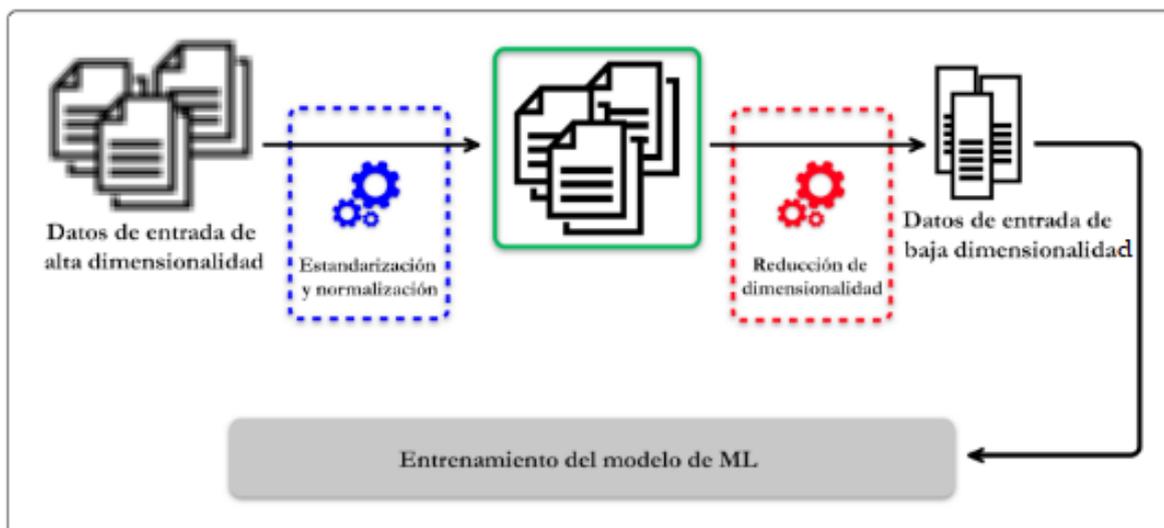


Figura B.1: Proceso de tratamiento de datos de alta dimensionalidad
Figura de elaboración propia

B.2 - La maldición de la dimensionalidad

Muchos escenarios de ML suelen contener conjuntos de datos con cientos o incluso miles de atributos, especialmente en dominios que tratan con datos complejos, como imágenes o flujos de video. Entonces se podría intuir que dichos escenarios otorgan mayor información y expresividad de datos, tendiendo a incrementar el rendimiento del modelo pero justamente acontece lo opuesto, es decir, el rendimiento y precisión de los modelos generados se degrada sensiblemente conforme la **dimensionalidad** aumenta.

El problema es que la densidad de la muestra disminuye exponencialmente respecto del aumento de la dimensionalidad, en otras palabras, el volumen del espacio crece tan rápidamente que los datos, al no aumentar del mismo modo, no pueden acompañar ese ritmo y por lo tanto, aumenta su dispersión. Por consiguiente, se requiere que el incremento dimensional sea correspondido con un aumento exponencial del número de instancias de entrenamiento para evitar que los modelos sobreajusten y reduzcan su rendimiento (ver **figura B.2**). [47]

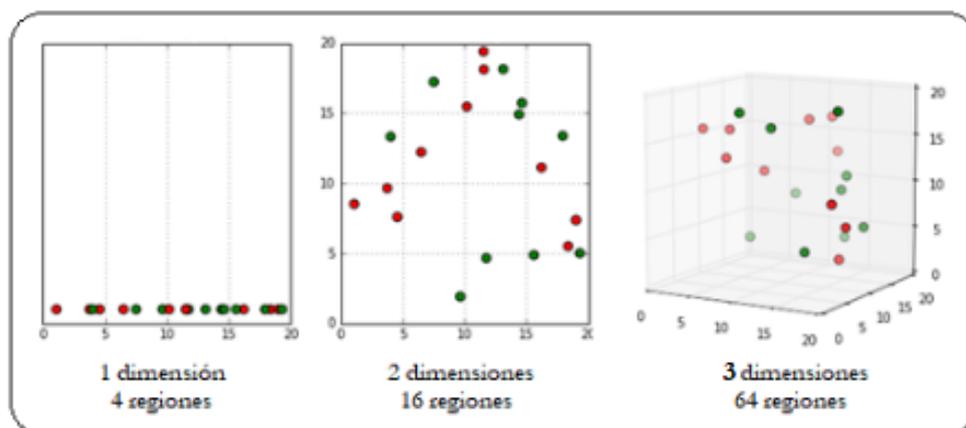


Figura B.2: Relación entre dimensionalidad y dispersión de los datos
Figura adaptada de **Curse of Dimensionality - DeepAI**

Por otro lado, las características adicionales actuarán como ruido para el modelo, que a su vez demorará más tiempo en completar su entrenamiento y deberá asignar mayor cantidad de recursos para el manejo de estos atributos innecesarios. En el mismo sentido, su estructura (grado de libertad) aumentará y tenderá a volverse más dependiente de los datos de entrenamiento (sobreajuste), disminuyendo su rendimiento predictivo sobre los datos de prueba reales. Adicionalmente, la varianza de los parámetros estimados tiende a incrementarse mientras que el sesgo de estimación y tamaño del conjunto de entrenamiento permanecen constantes. En resumen, estos factores ligados al aumento dimensional laceran de forma considerable el rendimiento de los modelos de ML generados, situación que en muchas ocasiones se designa como la **maldición de la dimensionalidad (CoD)**, la **figura B.3** mostrada a continuación ilustra lo explicado, de forma que para cualquier conjunto de datos existe una dimensionalidad óptima que posibilita el mayor rendimiento de los modelos y sobrepasada la misma, el mismo puede descender abruptamente. [45]

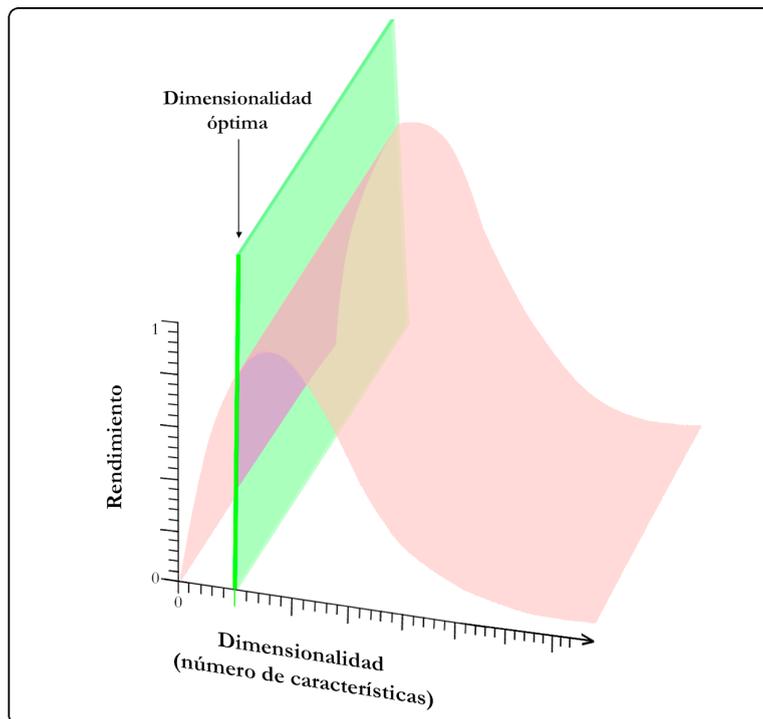


Figura B.3: Relación entre dimensionalidad y rendimiento en modelos de ML
Figura adaptada de **Confronting the Curse of Dimensionality- ODSC**

El proceso de reducción de dimensionalidad representa una etapa fundamental de preprocesamiento sobre aquellos conjuntos de datos de alta dimensionalidad. Su fin es la proyección de los datos originales, sobrecargados con información redundante, irrelevante o correlacionada, sobre un subespacio de dimensión reducida que preserve su esencia informativa, sin pérdida de variabilidad, interpretabilidad y expresividad, acotando el grado de libertad de los modelos para que alcancen mayor generalización, eficiencia y robustez.

B.3 - Técnicas y metodologías

Si bien existe una extensa gama de técnicas y metodologías de **DR**, área que ha cobrado importancia dentro del ML, motivando la dedicación de este anexo; no es el objetivo de esta sección recorrerlas todas ni profundizar excesivamente sobre cada una, sino exponer los conceptos y metodologías más sobresalientes.

En una visión general, la reducción de dimensionalidad puede subclasificarse en dos tipos de técnicas:

1. La **selección de características (FS)**, un proceso manual o automatizado de identificación de aquellas características más relevantes y la supresión de las que resultan improductivas (por redundancia, falta de expresividad, irrelevancia, alta correlación, etc.) para el modelo objetivo. Un problema de optimización que implica la extracción del subconjunto óptimo de características originales; de simple implementación pero con eventual pérdida informativa.
2. La **extracción o eliminación de características (FE)**, un proceso manual de generación de nuevas características mediante transformaciones del espacio de entrada original sobre un subespacio de baja dimensionalidad conservando la información relevante para el modelo.

Entre los algoritmos de selección (**FSA**s) tenemos:

- **Envoltorio**, consiste en la creación y evaluación de diferentes modelos sobre distintos subconjuntos de características de entrada a fin de hallar la combinación óptima que maximice el desempeño. Los métodos asociados son computacionalmente intensivos y de elevado costo, siendo únicamente aplicable con pocos atributos. Entre ellos, tenemos:
 - **Selección secuencial de características hacia adelante (SFS)**, método iterativo que añade progresivamente la característica que maximice el rendimiento hasta que una adición no represente mejora.
 - **Eliminación secuencial de características hacia atrás (SBE)**, método iterativo que inicia con todas las características y en cada iteración elimina la que produce el menor incremento de la tasa de error.
 - **Proporción de valores perdidos (MVD)**, elimina variables cuya proporción de valores inexistentes (nulos) supere un determinado umbral; considerando que es poco probable que características con abundantes datos faltantes contengan información significativa.
 - **Filtro de baja varianza (LVF)**, elimina aquellas variables cuya variabilidad esté por debajo de determinado umbral; considerándolas como de baja utilidad para el modelo objetivo.
 - **Filtro de alta correlación (HCF)**, unifica variables por pares, cuyo índice de correlación (Pearson o chi-cuadrado) supere determinado umbral; considerando que características muy correlacionadas mantienen información similar y dependiente.
 - **Bosques aleatorios (RF)**, determina características más relevantes por entrenamiento arbóreo, considerando que las usadas frecuentemente en la división, ameritan conservarse.

- **Filtro**, aplican técnicas estadísticas para ponderar el grado de relación entre cada variable de entrada y destino, usando esos puntajes como filtro selectivo del subconjunto de características más predictivas. No eliminan la multicolinealidad pero son de rápida ejecución. Entre ellos, tenemos:
 - **Correlación de Pearson**, cuantifica la dependencia lineal entre dos variables continuas.
 - **Análisis de varianza (ANOVA)**, similar a LDA sólo que opera mediante una o más funciones independientes categóricas y una función dependiente continua.
 - **Chi-Cuadrado**, prueba estadística que se aplica a características categóricas para evaluar la probabilidad de correlación entre ellos mediante una distribución de frecuencia.
- **Incrustación o integrados**, combina las cualidades de los métodos de envoltorio y filtro descriptos, mediante algoritmos que tienen incorporados sus propios métodos de selección de características. Entre los métodos correspondientes, tenemos:
 - **Regresión RIDGE**, originalmente propuesta para eludir los efectos adversos del problema de la colinealidad en un modelo lineal estimado por mínimos cuadrados. Incluye términos de penalización en la función objetivo para reducir el sobreajuste pero contrae los coeficientes hacia cero sin conseguir su nulidad, sin producir la selección de variables. [54]
 - **Regresión LASSO**, método de regresión lineal regularizada que efectúa la selección y regularización de características a fin de mejorar la interpretabilidad y reducir la variabilidad de estimaciones. Similar a RIDGE, con la mejora de admitir estimaciones nulas, realizando una especie de selección de variables en forma continua, debido a la norma L1. [54]

Entre los algoritmos de extracción (**FEAs**) tenemos métodos de transformaciones lineales y múltiples:

- **Lineales**, aplican transformaciones lineales para lograr la **DR**, algunos de los principales son:
 - **Análisis de componentes principales (PCA)**, método estadístico no supervisado de transformación ortogonal hacia un espacio de baja dimensión (componentes principales) minimizando la redundancia (covarianza) y maximizando la información (varianza).
 - **Análisis de componentes independientes (ICA)**, método que intenta hallar una representación que minimice la dependencia estadística de componentes representativos.
 - **Análisis discriminante lineal (LDA)**, técnica supervisada que proyecta los datos a fin de maximizar la separabilidad de clases.
- **No lineales o múltiples**, logran la **DR** mediante la aplicación de transformaciones sobre datos que no residen sobre un espacio lineal. Algunos de los principales métodos son:
 - **Escala multidimensional (MDS)**, analiza el grado de similitud de los datos a partir de la distancia Euclidiana del espacio geométrico proyectadas en una dimensión inferior.
 - **Mapeo de características isométricas (ISOMAP)**, proyecta los datos sobre una dimensión inferior preservando la distancia geodésica.

- **Incrustación de vecinos estocásticos distribuidos en t (t-SNE)**, calcula la probabilidad de relación entre pares de puntos en el espacio de alta dimensionalidad para luego seleccionar una inserción de baja dimensionalidad que produzca una distribución similar.
- **Análisis de discriminante generalizado (GDA)**, brinda un mapeo de vectores de entrada en el espacio original de alta dimensionalidad buscando una proyección de características de baja dimensionalidad maximizando dispersión entre clases y en cada clase.
- **Auto-codificador neural (NA)**, consiste en la utilización de autoencoders o codificadores automáticos (un tipo de ANN) para el aprendizaje de codificaciones y funciones de DR.

La principal diferencia entre las técnicas selección (FS) y extracción (FE) de características es que mientras FS efectúa la reducción gracias a la selección de un subconjunto de características, FE calcula una transformación de las originales, creando otras más significativas. El siguiente diagrama (**figura B.4**) resume las metodologías descriptas.

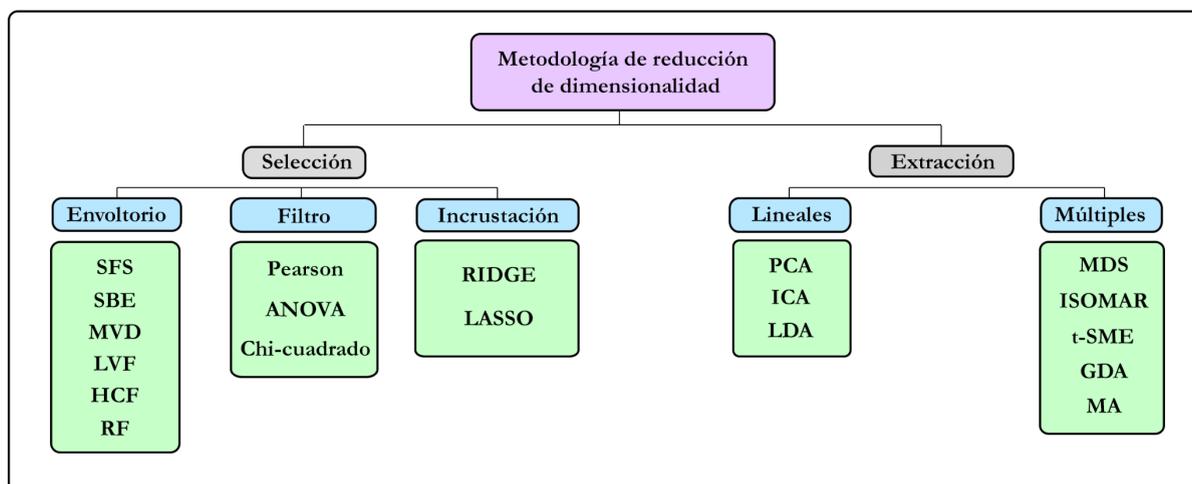


Figura B.4: Esquema de métodos de reducción de dimensionalidad
Figura de elaboración propia

C - Aplicaciones del aprendizaje por transferencia

En este anexo, se explorarán los más sobresalientes desarrollos y aplicaciones en el área del **TL** y se analizará de qué forma su avance ha contribuido a la mejora sustancial de las aplicaciones de **ML** existentes. El objetivo de esta sección no es realizar un exhaustivo análisis de cada desarrollo, sino que apunta a recorrer los principales trabajos y remarcar sus aspectos más innovadores a fin que el lector pueda mensurar del potencial de cada uno y la amplitud de ramificaciones aplicativas de la cuál goza el TL. Sumado a lo anterior, la presente sección nos permitirá anticiparnos a futuros requerimientos, delinear las distintas ramas de investigación así como las herramientas que surgirán en el corto plazo. Se recomienda enfáticamente al lector interesado profundizar en la investigación a partir de las referencias disponibles en cada trabajo. [22]

C.1 - Principales aplicaciones y desarrollos

En los últimos años, han habido grandes avances en redes neuronales profundas (**DNNs**) entrenadas con gran cantidad de datos etiquetados pero los modelos aún carecen de la habilidad de generalizar sobre las diferentes condiciones que se presentan en fases de entrenamiento; así como la capacidad de almacenar y transferir el conocimiento disponible sobre nuevos escenarios y condiciones, espíritu potenciador del TL.

C.1.1 - Agrupamiento y clasificación de texto

En la actualidad gran proporción del enorme caudal de información presente en internet existe como documentos textuales, pero debido a la migración de servicios tradicionales impresos al ámbito digital, han surgido técnicas automáticas de agrupamiento y clasificación con aplicaciones en gestión de contenido, búsqueda contextual, minería de opiniones y sentimientos, reseñas de productos, filtrado de correo indeseado, entre otros. En los últimos años se han aplicado exitosamente técnicas de TL sobre diferentes aplicaciones del mundo real, un ejemplo sobre el dominio textual es el trabajo titulado “**Construyendo información precedente mediante TL**”, que presenta un algoritmo de transferencia que construye un entorno de información sobre parámetros, que captura el conocimiento útil del dominio mediante dependencias subyacentes que permitan utilizar problemas similares para estimar la covarianza de parámetros que combinadas permiten aplicar una base de conocimiento que decreta el error en el rango del 20-40 %. [52]

Por su parte, el trabajo titulado “**Clasificación basada en co-agrupamiento para documentos fuera del dominio**”, donde el objetivo es transferir el conocimiento aprendido desde un dominio con datos etiquetados y distribución de probabilidad conocida, hacia otro diferente sin datos etiquetados mediante un algoritmo de clasificación basada en co-agrupamiento (COCC) utilizado como puente para propagar la estructura y el conocimiento entre dominios. Obteniendo un clasificador de documentos de alta calidad que supera tanto algoritmos de clasificación supervisados como semi-supervisados. Una línea de investigación futura se enfoca en la escalabilidad del agrupamiento y la configuración automática de los parámetros. [53]

C.1.2 - Agrupamiento y clasificación de imágenes

Esta área puede ser considerada como complementaria a la descrita previamente y que, entre otras aplicaciones, permite el análisis y procesamiento automático de las mismas para la mejora de aplicaciones de gestión forestal, estimación de capas de nieve, geología, ciberseguridad, reconocimiento facial y de objetos, conducción autónoma, diagnóstico de enfermedades y cuidado de la salud en general, etc.

El trabajo titulado **“Aprendizaje traducido: TL mediante diferentes espacios de características”**, que plantea un método considerado como una instancia de TL que consiste en transferir el conocimiento de los datos entre diferentes espacios de características. Para ello intenta construir un “puente”, mediante una ruta de vinculación utilizando cadenas de Markov que sirvan de nexo desde el espacio de características origen y objetivo, para la transferencia del conocimiento. Este trabajo puede ser utilizado para mejorar el aprendizaje de tareas muy diferentes, por ejemplo, disponiendo de datos etiquetados de clasificación textual (abundantes en la web), construir un clasificador de imágenes o la clasificación entre diferentes lenguajes. [25]

Por su parte, el trabajo titulado **“Adaptación de dominios desde múltiples fuentes mediante clasificadores auxiliares”**, propone un método para la adaptación de múltiples dominios (MDA) que permite el aprendizaje de una función de decisión robusta para la predicción de etiquetas mediante el aprovechamiento de un conjunto de clasificadores pre calculados desde múltiples dominios de origen o auxiliares. Este trabajo representa un avance significativo ya que MDA supera notablemente a otros métodos de adaptación de dominio existentes en un área cuya aplicación real considera la clasificación sentimental para reseñas de productos, la categorización textual, localización WiFi y detección de conceptos de video. [26]

Por su parte, el trabajo titulado **“TL heterogéneo para el agrupamiento de imágenes mediante la web social”**, presenta un nuevo escenario para mejorar el rendimiento del aprendizaje utilizando datos pertenecientes a diferentes espacios de características sin proporcionar una correspondencia entre instancias; ilustrando de qué forma la clasificación de imágenes puede también beneficiarse del **HetTL**. [27]

El trabajo titulado **“Mejorando la precisión de SVM mediante entrenamiento con fuentes de datos adicionales”**, se enmarca en la técnica de **ITL** y el enfoque de transferencia de instancias. Donde se propone la utilización de una fuente auxiliar de datos de entrada más abundantes pero de menor calidad que los de entrenamiento y prueba; demostrando que ante conjuntos de datos de entrenamiento muy reducidos es factible utilizar datos auxiliares para incrementar la precisión, reduciendo el error en un rango de 20-48 %. [50]

Finalmente el trabajo titulado **“Un estudio comparativo de métodos para TTL”**, estudia los casos más desafiantes de aprendizaje no supervisado del dominio de reconocimiento de entidades nombradas; que demuestra que incluso una pequeña cantidad de datos etiquetados con la distribución objetivo resulta beneficiosa para mejorar el rendimiento siendo incluso mejor que disponer de abundantes datos sobre D_s . [51]

C.1.3 - Aprendizaje reforzado

Entre las aplicaciones de **RL** asistido por técnicas de aprendizaje por transferencia encontramos diferentes dominios que han sido beneficiados, entre los cuales se encuentran: el aprendizaje robótico, ámbito de juegos, sistemas de **NLP**, informática de la salud, y distintas aplicaciones de la vida real, como sistemas de transporte, financieros, energéticos, entre otros. [41]

El trabajo titulado “**Modelado autónomo: transferencia de conocimiento en RL**”, introduce el modelado de recompensas donde un agente utiliza la experiencia previa de una secuencia de tareas para aprender un predictor de recompensas intermedias portable que sirve como acelerador de aprendizaje en futuras tareas relacionadas. De esta forma los agentes aprenden a estimar sus propias recompensas para nuevos estados, lo cuál produce una mejora sustancial del rendimiento en la resolución de tareas novedosas, incluso luego de breves períodos de entrenamiento. Este trabajo propone mejoras sobre sistemas RL (como robots en zonas de desastre o asistencia de invidentes) en cuanto a eficiencia y tiempos de entrenamiento. [24]

El trabajo titulado “**Transferencia entre dominios para RL**”, citado en el capítulo de **estado del arte**, completa una serie de trabajos relevantes de TL bajo el campo del RL, donde vimos de que forma la transferencia entre dominios puede ser notablemente mejorada por la técnica de transferencia. Destacando la importancia de definir una tarea de origen cuyo conocimiento sea beneficioso para aprender otra de interés, un ejemplo de esto se da cuando se intenta aprovechar el conocimiento obtenido a partir de un motor de videojuegos de carreras, como asistente del aprendizaje de conducción autónoma.

C.1.4 - Clasificación sentimental

El creciente caudal de información en internet se ha incrementado con el surgimiento de reseñas de los consumidores, haciendo que el análisis y utilización de esta información sea tan valiosa como el mismo producto o servicio ofrecido y siendo considerada de vital importancia para analistas de mercado, diseñadores, inversionistas, publicistas, entre otros. Como resultado de esto, los sistemas de ML que utilizan la clasificación sentimental han cobrado protagonismo y el TL promete llevarlos al siguiente paso, si de precisión y eficiencia se trata.

El trabajo titulado “**Adaptación de dominios para la clasificación de sentimientos**”, citado la **sección 3.2.3** del capítulo de **estado del arte**, representa la posibilidad de lograr soluciones superadoras en relación a la adaptación de dominios, mostrando que una reducida cantidad de datos etiquetados en el dominio objetivo resultan de gran ayuda para corregir posibles desajustes de correspondencia estructural. [9]

Por su parte el trabajo titulado “**Clasificación sentimental entre dominios usando un método en dos etapas**”, brinda un enfoque en dos etapas donde la primera identifica la correspondencia entre diferentes características de ambos dominios involucrados, utilizando palabras similares de diferentes dominios para conectar características específicas de cada uno permitiendo que características de ambos dominios se pongan en correspondencia para concretar la transferencia de conocimiento (tarea desempeñada por el método **TransferPLSA**) y la extracción de ese conocimiento compartido; mientras que la segunda etapa utiliza un clasificador entrenado con instancias etiquetadas de D_s para seleccionar algunos ejemplos informativos de D_T y luego reentrenarse con los ejemplos seleccionados, adaptándose a D_T . [34]

C.1.5 - Filtrado colaborativo

Como se ha mencionado previamente, el volumen de información disponible en internet aumenta a un ritmo asombroso y esa situación torna al problema de la búsqueda y acceso a dicha información relevante como desafiante e imprescindible; es aquí donde los sistemas de recomendaciones y donde particularmente el enfoque de filtrado colaborativo (**CF**) desempeñan su función.

El trabajo titulado “**Transferencia de aprendizaje para el filtrado colaborativo mediante un modelo generativo de matriz de calificaciones**”, demuestra que el modelo propuesto (RMGM) puede utilizarse para compartir el conocimiento y superar el problema de escasez de datos sobre ciertas tareas individuales en sistemas de recomendaciones de **CF** mediante el agrupamiento de múltiples matrices de calificaciones pertenecientes a dominios relacionados para la transferencia e intercambio de conocimiento. [35]

Por su parte, el trabajo titulado “**Transferencia de aprendizaje en filtrado colaborativo para reducción de escasez**”, intenta resolver uno de los principales problemas del contexto de sistemas de recomendaciones, el de la escasez o dispersión de datos, especialmente ante nuevos usuarios e items. Para ello se propone un novedoso método conocido como **sistema coordinado de transferencia (CST)** donde el conocimiento se integra desde fuentes de datos auxiliares. [36]

Finalmente, el trabajo titulado “**Filtrado colaborativo multidominio**”, afronta el problema del filtrado colaborativo multidominio (**MCF**) mediante la proposición de un **marco de factorización probabilística de matrices (PMF)** para modelar el problema de clasificación en cada dominio y permitir la transferencia de conocimiento de forma adaptativa mediante el conocimiento de la correlación entre ellos. [37]

C.1.6 - Estimación de ubicación basada en sensores

Con el vertiginoso avance que ha experimentado internet y la creciente conexión de objetos a la red, en lo que ha venido a denominarse como **internet de las cosas (IoT)**, ha emergido el campo de investigación de la localización y seguimiento de dispositivos en ambientes interiores, un problema sujeto a dificultades de dinamismo que hacen de su resolución algo tan atractivo como dificultoso. En parte, este tema fue estudiado en el trabajo analizado por la **sección 3.2.2**, se recomienda volver a dicha sección de ser necesario.

El trabajo titulado “**Transferencia de modelos de localización a través del espacio**”, propone un enfoque donde se comparten los datos de diferentes áreas en un entorno como puente de propagación del conocimiento del entorno completo, utilizando ese conocimiento transferido para reducir esfuerzo de calibración, al mismo tiempo que los datos etiquetados requeridos para la construcción del modelo de localización. A tales fines, se extrae el conocimiento del dominio del entorno de datos etiquetados para incorporarlo sobre un modelo que propague etiquetas sobre los datos sin etiquetar que fueron recopilados en el resto del entorno. [39]

Por último, el trabajo titulado “**TL mediante reducción de dimensionalidad**”, analizado en profundidad en la **sección 3.8 del capítulo de estado del arte** de la presente tesina, propone la utilización de un novedoso algoritmo (**MMDE**) que se vale de un espacio de características latente de baja dimensionalidad sobre el cual es posible aplicar algoritmos de aprendizaje estándar para entrenar modelos de clasificación o regresión. Para ello se sustenta en un nuevo método de reducción de dimensionalidad que minimiza la distancia existente entre las distribuciones de probabilidad de ambos dominios, resultando exitoso para el problema de localización en interiores mediante señales WiFi (**WILP**). [40]

C.1.7 - Aprendizaje de métricas

La mayoría de los enfoques de ML existentes, requieren contar con alguna medida de distancia entre los puntos de datos, contando con métricas estándar como euclidiana, bloque ciudadano, coseno, entre tantas otras. El aprendizaje de métricas de distancia tiene por objetivo el aprendizaje de las distancias presentes en los datos, que a su vez suelen agruparse en tres clases principales: análisis convexo, análisis matricial y teoría de la información.

El trabajo titulado “**Transfer Metric Learning by Learning Task Relationship**”, estudia el aprendizaje de métricas en torno al TL basado en el método **aprendizaje métrico de distancia regularizada (RDML)** y propone una extensión denominada **aprendizaje de métricas de transferencia (TML)** como mitigador del problema de deficiencia de datos etiquetados mediante la explotación de conocimiento útil desde varias tareas de origen independientes de igual importancia (se encuentra relacionado con **MTL**) a fin de mejorar el rendimiento sobre otra tarea objetivo determinada. Para el aprendizaje de las métricas de distancia de múltiples tareas de origen y sus relaciones con la tarea objetivo, este proceso es formulado como un problema de optimización convexo, modelando sus relaciones mediante una **matriz de covarianza** para finalmente adaptarla al escenario de transferencia en cuestión. [43]

Glosario

A

AdaBoost

Algoritmo basado en impulso o aceleración, desarrollado por Freund y Schapire en el año 1997, que representa un método general para incrementar la precisión de cualquier algoritmo de aprendizaje determinado.

Adaptación de dominios (DA, Domain Adaptation)

Subcampo del TL que trata con problemas que surgen al intentar obtener modelos de alta precisión donde difiere la distribución de los datos origen y destino; aprendiendo una función predictiva que minimice la divergencia entre ellos.

Agente de aprendizaje (LA, Learning Agent)

Herramienta del campo de la IA que posee la capacidad de aprender a partir de un conocimiento básico pero incrementándose y adaptándolo de forma autónoma mediante el aprendizaje de experiencias previas.

Agrupamiento (Clustering)

Método de aprendizaje no supervisado que divide los datos en grupos a partir del análisis de similitudes y diferencias. Entre los algoritmos más conocidos están: K-means, DbScan, Optics, Cure, Birch, Clarans, Sting y Clique, entre otros.

Algoritmo de árbol de decisión (DTa, Decision Tree algorithm)

Técnica para la aproximación de la función objetivo que representa la función aprendida en forma de árbol de decisión a partir de los valores de características, clasificando las instancias iterativamente desde la raíz hasta alguna de las hojas. Entre los algoritmos más conocidos se encuentran: Cart, Id3, Chaid, Chi cuadrado **C4.5**, C5.0 y **M5**, entre otros.

Algoritmo de agrupamiento autodidacta (StC, Self-taught Clustering)

Instancia correspondiente a una tarea de UTL que intenta agrupar una colección de datos destino no etiquetados, con la ayuda de gran cantidad de datos auxiliares, también sin etiquetar.

Algoritmo de agrupamiento de K-medias (K-means algorithm)

Popular algoritmo de ML no supervisado para el análisis de agrupamientos, cuyo objetivo es particionar N observaciones en K grupos a partir de su media, definida como el centro del mismo.

Algoritmo de agrupamiento de K-medias discriminativo (DisK-means algorithm)

Algoritmo eficiente de análisis de datos masivos que supera muchas deficiencias de K-medias inherentes a su naturaleza secuencial al aplicarlo sobre grandes conjuntos de datos.

Algoritmo de análisis de componentes principales (PCA, Principal Component Analysis algorithm)

Algoritmo de reducción de dimensionalidad empleado mayormente en el análisis de datos y construcción de modelos predictivos mediante transformaciones ortogonales que convierten variables correlacionadas en un conjunto lineal de variables no correlacionadas, denominadas componentes principales.

Algoritmo de análisis discriminativo transferido (TDA, Transferred Discriminative Analysis algorithm)

Método propuesto como solución al problema **TDR** que extrae información discriminativa de los datos etiquetados y los transfiere a una reducción de dimensionalidad no supervisada para una revisión iterativa.

Algoritmo de aprendizaje selectivo (SLA, Selective Learning Algorithm)

Algoritmo desarrollado para la resolución de problemas del tipo **DDTL**, que consiste de un proceso iterativo de utilización de instancias de dominios intermedios para obtener un clasificador de alta precisión.

Algoritmo de arroj (WNNa, Winnow algorithm)

Técnica de clasificación lineal para ejemplos etiquetados, que escala y se comporta bien con datos de alta dimensionalidad y fue diseñado para realizar una separación eficiente de atributos relevantes e irrelevantes.

Algoritmo de Bayes o Bayesiano (BA, Bayes Algorithm)

Técnica de clasificación que utiliza el teorema de probabilidad de Bayes, corresponde a la categoría de clasificadores probabilísticos y especialmente indicado cuando la dimensionalidad de las entradas es alta.

Algoritmo de bosque aleatorio (RFa, Random Forest algorithm)

Método de aprendizaje conjunto utilizado en problemas de clasificación y regresión que mediante el enfoque de “embolsado” crea un grupo de árboles de decisión que luego se combinan para formar árboles de decisión final. Consta de dos etapas: la creación del bosque aleatorio y la predicción a partir del bosque creado.

Algoritmo de codificación dispersa (SCA, Sparse Coding Algorithm)

Clase de algoritmos para hallar representaciones sucintas de estímulos a partir de datos de entrada sin etiquetar, descubriendo funciones que capturen características de nivel superior en los datos.

Algoritmo de mapeo convergencia (MCa, Mapping Convergence algorithm)

Algoritmo utilizado para alcanzar alta precisión de clasificación y consiste de una etapa de mapeo seguida de una de convergencia, de allí su nombre, y supera en precisión a la mayoría de los clasificadores tradicionales SVMs.

Algoritmo de transferencia por mapeo y revisión automática

(TAMaR, Transfer Automatic Mapping and Revision algorithm)

Método de TL que crea un mapeo directo entre los predicados de diferentes dominios.

Algoritmo FORTE (FORTE, First-Order Revision of Theories from Examples)

Algoritmo de refinamiento de cláusulas Horn mediante la integración de técnicas de revisión en un conjunto coherente compuesto por un marco de trabajo guiado por una heurística global. Es utilizado en varios dominios, incluyendo la programación lógica y el modelado cualitativo.

Algoritmo de K vecinos más cercanos (KNNa, K-Nearest Neighbours algorithm)

Método de aprendizaje supervisado basado en instancias, no paramétrico y muy utilizado en problemas de clasificación y regresión, que identifica y clasifica la clase de cada instancia en base a la proximidad con sus vecinos.

Aprendizaje (Learning)

Proceso a través del cual se adquieren y refinan habilidades, destrezas, conocimientos y conductas como resultado del proceso de estudio, aplicación de experiencias, instrucción, razonamiento y la observación.

Aprendizaje autodidacta o Auto-aprendizaje (StL, Self-taught Learning)

Paradigma de ML introducido por los investigadores de la universidad de Stanford en 2007, que procesa un conjunto de datos sin etiquetar para utilizarlos en el incremento de rendimiento de una determinada tarea supervisada.

Aprendizaje automático o automatizado (ML, Machine Learning)

Subcampo IA que estudia algoritmos computacionales y desarrolla técnicas que mejoran su rendimiento gracias al conocimiento incorporado de experiencias. Siendo capaces de aprender autónomamente a partir de un conjunto de datos.

Aprendizaje de correspondencia estructural (SCL, Structural Correspondence Learning)

Técnica novedosa de aprendizaje, perteneciente al área del aprendizaje y clasificación de sentimientos.

Aprendizaje de tareas múltiples o multi-tarea (MTL, Multi-Task Learning)

Enfoque de ITL que incrementa la generalización utilizando información de señales de entrenamiento en tareas relacionadas como sesgo inductivo. Aprendiendo múltiples tareas en paralelo mientras utiliza la representación compartida.

Aprendizaje por transferencia (TL, Transfer Learning)

Área de investigación activa del ML y técnica de aprendizaje que apunta a la obtención y utilización del conocimiento adquirido para la resolución eficiente de tareas o sobre dominios diferentes pero relacionados al original. Se puede subclasificar en las técnicas: inductiva (**ITL**), transductiva (**TTL**), no supervisada (**UTL**) y reforzada (**RTL**).

Aprendizaje por transferencia inductivo (ITL, Inductive Transfer Learning)

Técnica de TL que consiste en un mecanismo de aprendizaje que mejora el desempeño sobre una determinada tarea actual luego de haber aprendido una habilidad diferente pero relacionada de una tarea anterior. Algunos datos etiquetados en el dominio destino son requeridos para inducir el modelo predictivo objetivo.

Aprendizaje por transferencia heterogéneo (HetTL, Heterogeneous Transfer Learning)

Proceso de transferencia del conocimiento en situaciones donde tanto espacio de características como las distribuciones de probabilidades marginales de los dominios involucrados difieren. Además de adaptaciones entre distribuciones deben aplicarse adaptaciones del espacio de características, resultando más dificultoso que **HomTL**.

Aprendizaje por transferencia homogéneo (HomTL, Homogeneous Transfer Learning)

Proceso de transferencia del conocimiento en situaciones donde el espacio de características entre dominios es igual pero no así las distribuciones de probabilidad marginales asociadas. Por consiguiente requieren aplicarse técnicas de adaptación entre dominios para la corrección del **SSB** o del **CS**.

Aprendizaje por transferencia por refuerzo o reforzado (RTL, Reinforcement Transfer Learning)

Técnica de TL y área de investigación activa aplicada al paradigma del aprendizaje reforzado, que focaliza en el desarrollo de métodos para transferir el conocimiento desde un conjunto de tareas de origen hacia otro conjunto objetivo.

Aprendizaje por transferencia no supervisado (UTL, Unsupervised Transfer Learning)

Técnica de TL donde las tareas involucradas son diferentes y utiliza métodos no supervisados para aprender y resolver la tarea del dominio objetivo. Sin presencia de datos etiquetados en dominios origen o destino, en el entrenamiento.

Aprendizaje por transferencia transductivo (TTL, Transductive Transfer Learning)

Técnica de TL que asume diferentes dominios involucrados (ya sea distribuciones marginales de probabilidad o espacios de características) pero tareas de origen y destino son las mismas y además no existen datos etiquetados en el dominio destino disponibles durante el entrenamiento.

Aprendizaje profundo (DL, Deep Learning)

Subárea del ML que intenta modelar abstracciones de alto nivel mediante arquitecturas computacionales tensoriales no lineales (como ANNs o CNNs) compuesta de varios niveles o capas internas (ocultas). Admitiendo paradigmas de aprendizaje supervisado (SDL), semi-supervisado (SsDL) y no supervisado (UDL).

Análisis de discriminante Gaussiano (GDA, Gaussian Discriminant Analysis)

Modelo generativo para la clasificación de datos comúnmente utilizado cuando se asume que los datos pueden ser aproximados mediante una distribución normal multivariada o Gaussiana.

Análisis de discriminante lineal (LDA, Linear Discriminant Analysis)

Técnica de clasificación de datos y reducción de dimensionalidad que consiste de la generalización del discriminante lineal de Fisher. Se encuentra muy relacionado con el método **PCA**.

Área bajo la curva de recuperación de precisión (AUC, Area Under the precision-recall Curve)

Métrica de evaluación que considera los umbrales de clasificación posibles y mide hasta qué punto un modelo aprendido puede clasificar correctamente los átomos base de un conjunto de datos. Un área grande representa una alta recuperación (baja tasa de falsos negativos) y precisión (baja tasa de falsos positivos).

Aumentación de datos (Data+, Data Augmentation)

Conjunto de técnicas de incremento de datos de entrenamiento para el problema de carestía. Logra el incremento del tamaño y calidad de conjuntos de datos de entrenamiento y la consiguiente mejora de los modelos de ML. construidos.

Auto-codificador o autoencoder (Auto-Encoder)

Tipo específico de red neuronal hacia adelante no supervisada con una capa de entrada, una o más capas ocultas y una capa de salida que suele incluir dos procesos: codificación y decodificación. Trabaja comprimiendo la dimensionalidad de los datos de entrada y reconstruyendo la salida desde esa representación.

B

Base de conocimiento (KB, Knowledge Base)

Tipo especial de base de datos para la gestión del conocimiento, administrada por un motor de inferencia (KBMS) que aplica reglas para obtener respuestas. Pueden clasificarse en: legibles por máquinas y legibles por humanos.

Búsqueda del camino relacional (RPF, Relational Pathfinding)

Método general propuesto por Richards y Mooney en 1992, que ayuda a los sistemas de aprendizaje de primer orden a evitar máximos locales y también a atravesar mesetas locales.

C

Cambio covariante (CS, Covariate Shift)

Situación que acontece cuando existe cierta variación en la distribución de probabilidades asociada a los datos de entrada entre las fases de entrenamiento y prueba, resultando en un modelo predictivo erróneo y de pobre rendimiento.

D

Despliegue de varianza máxima (MVU, Maximum Variance Unfolding)

Método heurístico de reducción de dimensionalidad que produce una representación de baja dimensionalidad maximizando la varianza de los agrupamientos mientras conserva las distancias locales originales de cada agrupamiento.

Dimensionalidad (Dimensionality)

Cantidad de variables aleatorias (características o atributos de entrada) que contiene un determinado conjunto de datos, en una representación tabular de los datos, puede verse como la cantidad de columnas existentes.

Distribución de probabilidad (PD, Probability Distribution)

Función estadística que describe todos los valores posibles y probabilidades que una variable aleatoria admite en un rango determinado. Permitiendo asignar a cada suceso definido sobre la variable, un valor de probabilidad de ocurrencia.

Distribución de probabilidad condicional (CPD, Conditional Probability Distribution)

Función de asignación de probabilidad sobre eventos cuya ocurrencia depende de la ocurrencia de otro evento asociado. El teorema de Bayes es crucial en el análisis de este tipo de distribuciones.

Distribución de probabilidad marginal (MPD, Marginal Probability Distribution)

En teoría de probabilidades, referida como la distribución de probabilidad de un subconjunto de variables aleatorias. Su nombre se debe a la disposición de probabilidades conjuntas en forma tabular, se ubicarían en los márgenes de la tabla.

Dominio (Domain)

Ámbito o entorno de un sistema de ML representado por los datos de entrada y compuesto por dos componentes: un espacio de características y una distribución de probabilidad marginal asociada.

E

Error cuadrático medio (MSE, Mean Squared Error)

Estimador que mide el promedio de errores cuadrados, es decir, la diferencia entre el estimador y el valor real estimado. Su propósito es evaluar la calidad asociada a un predictor o estimador determinado.

Espacio de características (FS, Feature Space)

Espacio n-dimensional donde las variables de entrada conviven, en la representación usual tabular del conjunto de datos de entrada, se corresponde con el conjunto de columnas presentes y es determinante para la calidad de los modelos.

Espacio de etiquetas (LS, Label Space)

Conjunto de valores posibles asignables a cada instancia predicha por un modelo de ML. Es la imagen de la **función predictiva** ejecutada por el modelo y puede ser un espacio discreto multidimensional (clasificación) o continuo (regresión).

Estimación de densidad (DE, Density Estimate)

En probabilidades y estadísticas, se refiere como la construcción de la estimación, basada en los datos observados, de una función de probabilidad de densidad subyacente, es decir, la densidad de valores continuos en un rango de acuerdo a las muestras, implicando la relación entre las observaciones y sus probabilidades.

Extracción de características (FE, Feature Extraction)

Proceso de reducción de dimensionalidad por el cual un conjunto inicial de datos “crudos” (sin preprocesar), es reducido a grupos más manejables, minimizando los recursos necesarios para su tratamiento.

F

Filtrado colaborativo (CF, Collaborative Filtering)

Metodología de optimización aplicada a sistemas de recomendación y administración de grandes conjuntos de datos. Implementan algoritmos matemáticos potentes para el establecimiento de criterios de preferencias y filtrado de información donde la preferencia de un usuario sobre un ítem es predicha en base a otros usuarios con intereses similares.

Función predictiva (PF, Predictive Function)

En ML es una función generada a partir de la fase de entrenamiento de un modelo, que asigna a cada entrada un valor predictivo de salida y puede ser probada en fase de pruebas para medir la eficiencia y calidad predictiva del modelo.

G

Generalización (Generalization)

Habilidad de desempeño propia de un modelo de ML. Representa una característica deseable y medible, que implica no sólo el aprendizaje desde los datos de entrenamiento sino además una elevada precisión predictiva con datos de pruebas.

I

Independiente e idénticamente distribuido (IID, Independently and Identically Distributed)

Datos obtenidos de una distribución inmutable, en los que cada valor no depende de valores obtenidos previamente. Representan el combustible ideal del ML; son una construcción matemática útil pero difícil de hallar en el mundo real.

Instancia (Instance)

También denominada como “muestra” o “ejemplo”, se refiere a cada uno de los datos disponibles, compuesto por características descriptivas, para ser utilizados por los modelos de ML en fases de entrenamiento, validación y pruebas.

Inteligencia

En el ámbito del ML, capacidad computacional para aprender a realizar una determinada tarea para lo cual se requiere entrenamiento mediante la provisión de datos de entrada, sin necesidad de utilizar un lenguaje de programación imperativo para comunicarle a la máquina una serie de acciones a realizar.

M

Máquina de vectores de soporte (SVM, Support Vector Machine)

Conjunto de algoritmos de aprendizaje supervisado que pueden utilizarse en problemas tanto de clasificación como de regresión. Trabajan bajo el concepto de cálculo de márgenes, maximizando la distancia entre puntos cercanos de diferentes clases y el hiperplano de separación (margen), que puede ser rígido o flexible.

Máquina de vectores de soporte de margen flexible (SMSVM, Soft Margin Support Vector Machine)

Tipo de clasificador SVM al que se le permite cometer cierto número de errores para intentar conservar un margen de decisión lo más amplio posible.

Máquina de vectores de soporte transductiva (TSVM, Transductive Support Vector Machine)

Paradigma utilizado para el tratamiento de datos parcialmente etiquetados en el campo **SSL**, que busca la maximización del margen de separación entre datos etiquetados y sin etiquetar mediante regularización. Los SVMs tradicionales utilizan sólo datos etiquetados mientras que los TSVMs también hacen uso de datos sin etiquetar.

Marco de trabajo conjunto ponderado localmente (LWE, Locally Weighted Ensemble framework)

Algoritmo que incorpora la categorización de clasificadores e incrementa la precisión de clasificación a partir de la incorporación de información proveniente de diferentes modelos.

Marco de trabajo de promediado simple de modelo (SMA, Simple Model Averaging framework)

Marco que efectúa la combinación de múltiples modelos de TL donde los pesos son asignados dinámicamente de acuerdo al poder predictivo de cada modelo sobre cada ejemplo de prueba, esto permite integrar las ventajas de varios algoritmos de aprendizaje en un modelo unificado aplicable sobre un dominio diferente.

Matriz de covarianza (CM, Covariance Matrix)

En estadística y teoría de probabilidad, matriz cuadrada que contiene la covarianza de los elementos de un determinado vector.

Método de incrustación máxima de discrepancia media

(**MMDE, Maximum Mean Discrepancy Embedding**)

Método de reducción de dimensionalidad consistente en el aprendizaje de un espacio latente común a ambos dominios involucrados, minimizando el ruido y distancia entre distribuciones, para efectivizar el proceso de transferencia.

Método de información mutua (MIM, Mutual Information Method)

Algoritmo que mide la reducción de incertidumbre entre dos variables aleatorias a partir de un valor conocido de una de ellas, consistiendo de una medida de dependencia mutua entre dos variables aleatorias.

Método de información mutua de aprendizaje por correspondencia estructural

(SCL-MI, Structural Correspondence Learning Mutual Information)

Algoritmo de transferencia que selecciona un conjunto de características pivotes de alta frecuencia en ambos dominios y luego establece correspondencias con características de datos sin etiquetar y así predecir ocurrencias.

Método de transferencia de reglas (RT_m, Rule Transfer method)

Método de transferencia de aprendizaje reforzado que permite el bajo acoplamiento de representaciones internas entre agentes a partir del aprendizaje de reglas y su manipulación para optimizar la transferencia de conocimiento entre ellos.

Minería de datos (DM, Data Mining)

Campo de la estadística y la ciencia de computación referido a los procesos de descubrimiento de patrones y relaciones relevantes sobre grandes volúmenes de datos mediante técnicas analíticas para la toma de decisiones favorables.

Minimización del riesgo empírico (ERM, Empirical Risk Minimization)

Principio de la teoría de aprendizaje estadístico que define una familia de algoritmos de aprendizaje y es usado para brindar límites teóricos de desempeño para los modelos.

Modelo de aprendizaje automático (MLM, Machine Learning Model)

Artefacto computacional que ha sido entrenado para el reconocimiento de ciertos patrones a partir de datos de entrada proporcionados y que implementa un determinado algoritmo de aprendizaje para su razonamiento. Una vez entrenado, el modelo podrá realizar predicciones sobre datos previamente desconocidos.

Modelado predictivo (PM, Predictive Modeling)

Técnica estadística utilizada para realizar predicciones sobre el comportamiento futuro en base al análisis de datos.

Muestra insesgada (US, Unbiased Sample)

Muestra extraída y registrada mediante un método libre de sesgos. Cada muestra debe tener la misma probabilidad de selección, ser representativa de la población y lo suficientemente amplia como para generalizar los resultados obtenidos.

Muestreo de importancia (IS, Importance Sampling)

Poderosa técnica estadística utilizada para la estimación de la esperanza propia de una variable aleatoria bajo una distribución de probabilidad p desde muestras que presentan una distribución q diferente.

P

Paradigma (paradigm)

Modelo, enfoque o patrón de conocimiento que debe seguirse en determinada situación o para la resolución de determinados problemas y que es aceptado o sostenido por una disciplina o comunidad científica.

Paradigma de aprendizaje no supervisado (ULP, Unsupervised Learning Paradigm)

Paradigma de ML donde el proceso de aprendizaje se realiza independientemente de cualquier tipo de supervisión (datos de entrenamiento sin etiquetar) y consiste en el reconocimiento de patrones existentes en los datos para la derivación de reglas. Indicado para situaciones donde las categorías de los datos son desconocidas.

Paradigma de aprendizaje por refuerzo o reforzado (RLP, Reinforcement Learning Paradigm)

Paradigma de ML donde un agente de software toma decisiones en un entorno con el objetivo de maximizar la recompensa acumulativa. Los algoritmos pertenecientes a este paradigma deben explorar y descartar alternativas para obtener el resultado correcto sin disponer de sugerencias, únicamente la indicación de correctitud de la recompensa.

Paradigma de aprendizaje semi-supervisado (SsLP, Semi-supervised Learning Paradigm)

Paradigma de ML que aprovecha el poder de las técnicas de aprendizaje supervisado (con instancias etiquetadas) y no supervisado (sin instancias etiquetadas) ya que podrían darse situaciones donde sólo un subconjunto de datos se encuentren etiquetados, relajando la dependencia de instancias etiquetadas y reduciendo los costos asociados a ese proceso.

Paradigma de aprendizaje supervisado (SLP, Supervised Learning Paradigm)

Paradigma de ML donde el proceso de entrenamiento se encuentra bajo supervisión mediante la introducción de datos de entrada junto a la salida deseada que luego es comparada con la salida producida, generando una señal de error que permitirá el ajuste de los parámetros a fin de aumentar la precisión del modelo.

PCA+K-medias (PCA+K-means framework)

Marco de trabajo que plantea la combinación de técnicas PCA y K-medias para dimensionalidad transferida.

Pérdida Huber (HL, Huber Loss)

Función de pérdida utilizada en regresión robusta y que resulta ser menos sensible a valores muestrales atípicos de los datos que la función de pérdida de error cuadrático (MSE).

Pérdida promedio (AL, Average Loss)

Medida que indica el valor medio de error de un modelo para una serie de instancias, es decir, la distancia existente entre la predicción del modelo para determinada instancia de entrenamiento y su valor real o etiqueta. Es calculada en fases de entrenamiento y validación, cuanto menor es, mejor será el modelo evaluado (sin sobreajustar).

Precisión o exactitud (Accuracy)

Métrica de evaluación de modelos de clasificación que indican la proporción de predicciones que el modelo realizó correctamente, simbolizando la relación entre el número de predicciones correctas y el total de predicciones. Naturalmente cuando la precisión calculada tiende a 1 (o al 100%), mejor es considerado el modelo.

Probabilidad de registro condicional (CLL, Conditional Log-Likelihood)

Mide hasta qué punto un modelo predice de forma precisa probabilidades de átomos base de un conjunto de datos.

Probabilidad de registro pseudo ponderado (WPLL, Weighted Pseudo Log-Likelihood)

Aproximación de probabilidad y métrica para etapas de optimización de algoritmos estructurales y ponderación.

Problema de adaptación de dominios (DAP, Domain Adaptation Problem)

Problema de TL donde el objetivo es la aplicación de un modelo con buen desempeño desde una determinada distribución de origen sobre otra destino diferente (relacionada). Puede resolverse mediante enfoques supervisados, semi-supervisados o no supervisados.

Problema de localización en interiores basado en señales WiFi

(WILP, WiFi-based Indoor Localization Problem)

Compleja y muy importante tarea para el área de la AI y aplicaciones de cómputo ubicuo. Al ser una tarea altamente dependiente de cambios contextuales, la convierte en ideal para aplicar enfoques de TL para su resolución.

Proceso de entrenamiento (TP, Training Process)

Etapas o estadio de construcción de modelos de ML cuyo objetivo es mejorar gradualmente la capacidad predictiva del mismo mediante el análisis de tasas de error y precisión analizados para el consiguiente ajuste de hiper parámetros.

Proceso de decisión de Markov (MDP, Markov Decision Process)

Proceso de control estocástico, para entornos discretos y secuenciales, que proporciona un marco matemático de modelado para la toma de decisiones de un agente, en situaciones donde intervienen factores aleatorios y bajo su control.

Procesamiento de lenguaje natural (NLP, Natural Language Processing)

Campo de IA y la lingüística computacional que estudia las interacciones (aprendizaje, interpretación y manipulación) entre computadoras y el lenguaje humano.

Proceso Gaussiano (GP, Gaussian Process)

En teoría de probabilidad y estadística, proceso estocástico que genera muestras cuya combinación lineal poseerá una distribución normal (o Gaussiana).

Programación lógica inductiva (ILP, Inductive Logic Programming)

Colección de métodos de aprendizaje basados en la lógica y estadística, que son utilizados para generar bases de conocimiento consistentes en fórmulas de la lógica de primer orden que modelan un conjunto de datos determinado.

Punto de acceso (AP, Access Point)

Dispositivo de red que interconecta equipos de comunicación inalámbricos móviles, formando una red inalámbrica.

R

Regresor mínimo cuadrado regularizado (RLSR, Regularized Least Square Regressor)

Modelo de regresión que utiliza el método de mínimos cuadrados y una fase de regularización para restringir aún más la solución encontrada. Utilizados con frecuencia cuando el modelo aprendido posee un grado pobre de generalización.

Red lógica de Markov (MLN, Markov Logic Network)

Estructura de representación del conocimiento que combina la formalidad compacta y expresividad de la lógica de primer orden sobre dominios finitos con la flexibilidad de las probabilidades, para el aprendizaje de dominios relacionales.

Red neuronal artificial (ANN, Artificial Neural Network)

Modelo inspirado en el funcionamiento de su homólogo biológico humano donde la unidad básica de aprendizaje es la neurona, una célula nerviosa. Pueden clasificarse en: supervisadas (SNN), no supervisadas (UNN) y reforzadas (RNN).

Red neuronal convolucional (CNN, Convolutional Neural Network)

Tipo de ANN basada en un proceso de “convolución” donde se recorre una imagen de entrada aplicando un filtro y obteniéndose una nueva imagen resultante. Representan un salto de calidad en la clasificación de imágenes.

Reducción de dimensionalidad (DR, Dimensionality Reduction)

Metodología que intenta reducir el costo computacional mediante la reducción de dimensiones (cantidad de atributos) de los datos, eliminando datos redundantes e irrelevantes para mejorar la precisión de los resultados.

Reducción de dimensionalidad transferida (TDR, Transferred Dimensionality Reduction)

Problema donde se transfiere la información de una tarea relacionada desde clases previas conocidas, para dominios de clase sin etiquetar, intentando hallar el mejor subespacio discriminativo.

Regresión de vectores de soporte (SVR, Support Vector Regression)

Algoritmo de aprendizaje supervisado similar a SVM pero que utiliza un mapeo no lineal.

Regresión logística (LR, Logistic Regression)

Método que modela probabilidades de clasificación con dos posibles resultados y es extensión de la regresión lineal.

Regresor Laplaciano de mínimos cuadrados regularizado

(LapRLSR, Laplacian Regularized Least Square Regressor)

Tipo de regularización para modelos de regresión utilizando la distribución de Laplace o doble exponencial.

S

Selección de características (FS, Feature Selection)

Proceso de reducción de dimensionalidad mediante el cual se selecciona, manual o automáticamente, aquellas características que contribuyen en mayor medida a la predicción de resultados, aumentando la precisión de los modelos.

Sesgo de selección de la muestra (SSB, Sample Selection Bias)

Distorsión de selección introducida por el método de recolección. Predisposición hacia un resultado particular por sobre otro, situación que conduce a resultados no representativos y reduce la confianza de los datos.

Sobreajuste (Overfitting)

Situación que ocurre cuando la brecha entre el error en etapas de entrenamiento y prueba se torna muy grande. Interpretado como imposibilidad de generalización y memorización de datos de entrenamiento por parte del modelo.

Subajuste (Underfitting)

Situación que ocurre cuando un modelo de ML no obtiene valores de error lo suficientemente bajos sobre el conjunto de entrenamiento. Interpretado como su excesiva simplicidad al no representar la complejidad del problema.

SVM de margen suave (SVM-Light)

Subtipo de máquina de vectores de soporte que admite el tratamiento de datos no linealmente separables, incorporando un parámetro de regulación que determina el grado de tolerancia a violaciones sobre el margen de separación.

T

Tarea (Task)

Problemática planteada que presupone la construcción de modelos de ML para su resolución y abarca el proceso de predicción o inferencia efectuado en función de un problema y los datos disponibles.

TrAdaBoost

Marco de trabajo basado en el algoritmo **AdaBoost**, que produce mejoras significativas de clasificación gracias al uso de datos etiquetados sobre un dominio con diferente distribución, como datos de entrenamiento adicionales.

Transferencia entre dominios distantes (DDTL, Distant Domain Transfer Learning)

Categoría de problemas que implica la transferencia de aprendizaje entre dominios conceptualmente distantes, es decir, con un grado relacional extremadamente bajo donde la transferencia por fuerza bruta resulta perjudicial.

Transferencia positiva (PT, Positive Transfer)

Situación deseada que acontece cuando la aplicación de métodos de TL contribuyen al incremento del rendimiento del modelo. Su probabilidad de ocurrencia aumenta cuando las tareas se encuentran apropiadamente relacionadas.

Transferencia negativa (NT, Negative Transfer)

Situación indeseada que acontece cuando la aplicación de un método de TL contribuye a reducir el rendimiento, en lugar de incrementarlo. El grado relacional entre tareas y el enfoque adoptado determinan la probabilidad de que acontezca.

Referencias bibliográficas

1 - **A Survey on Transfer Learning**

Sinno Jialin Pan and Qiang Yang Fellow - IEEE, NEC Lab. - China - 2010

Fecha de acceso: 15/07/2020

2 - **Transfer Learning**

Lisa Torrey and Jude Shavlik - University of Wisconsin - Madison WI, USA - 2009

Fecha de acceso: 18/07/2020

3 - **Lifelong Machine Learning Systems: Beyond Learning Algorithms**

Daniel Silver and Qiang Yang and Linghao Li - AAAI - 2013

Fecha de acceso: 24/07/2020

4 - **Transfer Learning using Computational Intelligence: A Survey**

Jie Lu, Vahid Behbood and cols. - University Technology of Sydney, Australia - 2015

Fecha de acceso: 28/07/2020

5 - **Learning to Learn: Knowledge Consolidation and Transfer in Inductive Systems**

NIPS*95 Post-Conference Workshop - Marriott Hotel, Colorado, USA - 1995

Fecha de acceso: 29/07/2020

6 - **Lifelong Machine Learning: a paradigm for continuous learning**

Bing Liu - University of Illinois, Chicago, USA - 2016

Fecha de acceso: 02/08/2020

7 - **PEBL - Web Page Classification without Negative Examples**

H. Yu, J. Han, K. Chen-Chuan Chang, IEEE, China - 2004

Fecha de acceso: 05/08/2020

8 - **WILP: Transfer Learning for WiFi-based Indoor Localization**

Sinno Jialin Pan, Vincent Wenchen Zheng, Qiang Yang and Derek Hao Hu,

Hong Kong University of Science and Technology, China - 2008

Fecha de acceso: 09/08/2020

9 - **Domain Adaptation for Sentiment Classification**

John Blitzer, Mark Dredze and Fernando Pereira

Department of Computer and Information Science, University of Pennsylvania USA - 2007

Fecha de acceso: 10/08/2020

10 - **Cross-Domain Transfer for Reinforcement Learning**

Matthew E. Taylor and Peter Stone

Department of Computer Sciences, University of Austin Texas USA - 2007

Fecha de acceso: 12/08/2020

11 - **Self-taught Learning: Transfer Learning from Unlabeled Data**

Rajat Raina, Alexis Battle, Honglak Lee, Benjamin Packer and Andrew Y. Ng

Computer Science Department, Stanford University California USA - 2006

Fecha de acceso: 16/08/2020

- 12 - **Learning to Learn with the Informative Vector Machine**
Neil D. Lawrence and John C. Platt
Microsoft Research, Redmond Washington USA - 2004
Fecha de acceso: 16/08/2020
- 13 - **Mapping and Revising Markov Logic Networks for Transfer Learning**
Lilyana Mihalkova and Tuyen Huynh and Raymond J. Mooney
Department of Computer Sciences, The University of Texas at Austin, USA - 2007
Fecha de acceso: 18/08/2020
- 14 - **Transfer Learning by Mapping with Minimal Target Data**
Lilyana Mihalkova and Raymond J. Mooney
Department of Computer Sciences, The University of Texas at Austin, USA - 2008
Fecha de acceso: 24/08/2020
- 15 - **Knowledge Transfer Via Multiple Model Local Structure Mapping**
Jing Gao, Wei Fan, Jing Jiang and Jiawei Han
Institutional Knowledge at Singapore Management University, Singapore - 2008
Fecha de acceso: 13/09/2020
- 16 - **Learning and Evaluating Classifiers under Sample Selection Bias**
Bianca Zadrozny
IBM T.J. Watson Research Center, Yorktown Heights, Nueva York, USA - 2004
Fecha de acceso: 20/09/2020
- 17 - **Feature Selection for Dimensionality Reduction**
Dunja Mladenic
Jozef Stefan Institute, Ljubljana, Slovenia - 2006
Fecha de acceso: 20/09/2020
- 18 - **Analysis of Dimensionality Reduction - Techniques on Big Data**
G. Reddy, K. Lakshmana, R. Kaluri, D. Rajput, G. Srivastava and T. Barker,
IEEE, USA - 2020
Fecha de acceso: 26/09/2020
- 19 - **Boosting for Transfer Learning**
Wenyuan. Dai, Qiang Yang, Gui-Rong Xue and Yong Yu
Oregon, USA - 2007
Fecha de acceso: 29/09/2020
- 20 - **Transferred Dimensionality Reduction**
Zheng Wang, Yangqiu Song, and Changshui Zhang
TNList, Tsinghua University, Beijing, China - 2007
Fecha de acceso: 29/09/2020
- 21 - **A survey on Image Data Augmentation for Deep Learning**
Connor Shorten and Taghi Khoshfogaar
Journal of Big Data, USA - 2019
Fecha de acceso: 25/10/2020

- 22 - **Some Applications in Transfer Learning**
Fuzhen Zhuang and Ping Luo
17° ACM CIKM, Napa Valley, California, USA - 2008
Fecha de acceso: 27/10/2020
- 23 - **Using Advice to Transfer Knowledge Acquired in One RL Task to Another**
Lisa Torrey, Trevor Walker, Jude Shavlik and Richard Maclin
16° CML European Conference on Machine Learning, Porto, Portugal - 2005
Fecha de acceso: 28/10/2020
- 24 - **Autonomous Shaping: Knowledge Transfer in Reinforcement Learning**
George Konidaris and Andrew Barto
23° Annual International Conference on Machine Learning (ICML), Pennsylvania, USA - 2006
Fecha de acceso: 05/12/2020
- 25 - **Translated Learning: Transfer Learning across Different Feature Spaces**
Wenyuan Dai, Yuqiang Chen, Gui-rong Xue, Qiang Yang and Yong Yu
22° Neural Information Processing Systems (NIPS), Vancouver, Canada - 2008
Fecha de acceso: 06/12/2020
- 26 - **Domain Adaptation from Multiple Sources via Auxiliary Classifiers**
Lixin Duan, Ivor W. Tsang, Dong Xu and Tat-Seng Chua
26° International Conference on Machine Learning (ICML), Montreal, Canada - 2009
Fecha de acceso: 07/12/2020
- 27 - **Heterogeneous Transfer Learning for Image Clustering via the Social Web**
Qiang Yang, Yuqiang Chen, Gui-Rong Xue, Wenyuan Dai and Yong Yu
47° Annual Meeting of the Association for Computational Linguistics (ACL) - 2009
Fecha de acceso: 08/01/2021
- 28 - **A Comprehensive Survey on Transfer Learning**
Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong and Qing He
Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), China- 2019
Fecha de acceso: 09/01/2021
- 29 - **A survey on Transfer Learning**
Karl Weiss , Taghi M. Khoshgoftaar and DingDing Wang
Journal of Big Data, USA - 2016
Fecha de acceso: 10/01/2021
- 30 - **Deep Transfer via Second-Order Markov Logic**
Jesse Davis y Pedro Domingos
Department of Computer Science & Engineering, University of Washington, Seattle, USA - 2008
Fecha de acceso: 10/01/2021
- 31 - **Heterogeneous Transfer Learning for Image Clustering via the Social Web**
Qiang Yang, Yuqiang Chen, Gui-Rong Xue, Wenyuan Dai and Yong Yu
Hong Kong University of Science and Technology, Clearway Bay, Hong Kong, China - 2009
Fecha de acceso: 20/01/2021

- 32 - **Machine Learning from Theory to Algorithms: An Overview**
Jafar Alzubi, Anand Nayyar and Akshi Kumar
Journal of Physics Conference Series, USA - 2018
Fecha de acceso: 22/01/2021
- 33 - **Distant Domain Transfer Learning**
Ben Tan, Yu Zhang, Sinno Jialin Pan and Qiang Yang
Hong Kong University of Science and Technology, Hong Kong, China - 2017
Fecha de acceso: 23/01/2021
- 34 - **Cross-Domain Sentiment Classification Using a Two-Stage Method**
Kang Liu and Jun Zhao
Institute of Automation, Chinese Academy of Sciences, Beijing, China - 2009
Fecha de acceso: 27/01/2021
- 35 - **Transfer Learning for Collaborative Filtering via a Rating-Matrix Generative Model**
Bin Li, Qiang Yang and Xiangyang Xue
School of Computer Science, Fudan University, Shanghai, China - 2009
Fecha de acceso: 28/01/2021
- 36 - **Transfer Learning in Collaborative Filtering for Sparsity Reduction**
Weike Pan, Evan W. Xiang, Nathan N. Liu and Qiang Yang
Department of Computer Science and Engineering University of Science and Technology, Hong Kong, China - 2010
Fecha de acceso: 03/02/2021
- 37 - **Multi-Domain Collaborative Filtering**
Yu Zhang, Bin Cao and Dit-Yan Yeung
Department of Computer Science and Engineering, Hong Kong University of Science and Technology, China - 2010
Fecha de acceso: 11/02/2021
- 38 - **Transferring Knowledge from Another Domain for Learning Action Models**
Hankui Zhuo, Qiang Yang, Derek Hao Hu and Lei Li
Software Research Institute, Sun Yat-sen University, Hong Kong University of Science and Technology, China - 2010
Fecha de acceso: 15/02/2021
- 39 - **Transferring Localization Models Across Space**
Sinno Jialin Pan, Dou Shen, Qiang Yang and James T. Kwok
Hong Kong University of Science and Technology, Hong Kong, China - 2008
Fecha de acceso: 25/02/2021
- 40 - **Transfer Learning via Dimensionality Reduction**
Sinno Jialin Pan, James T. Kwok and Qiang Yang
Hong Kong University of Science and Technology, Hong Kong, China - 2008
Fecha de acceso: 06/03/2021
- 41 - **Transfer Learning in Deep Reinforcement Learning: A Survey**
Zhuangdi Zhu, Kaixiang Lin and Jiayu Zhou
Hong Kong University of Science and Technology, Hong Kong, China - 2020
Fecha de acceso: 09/03/2021

- 42 - **Knowledge transfer for cross domain learning to rank**
Depin Chen, Yan Xiong, Jun Yan, Gui-Rong Xue, Gang Wang and Zheng Chen
Hong Kong University of Science and Technology, Hong Kong, China - 2009
Fecha de acceso: 10/03/2021
- 43 - **Transfer Metric Learning by Learning Task Relationships**
Yu Zhang and Dit-Yan Yeung
Hong Kong University of Science and Technology, Hong Kong, China - 2009
Fecha de acceso: 10/03/2021
- 44 - **Transfer Learning for Reinforcement Learning Domains: A Survey**
M. Taylor and P. Stone
Journal of Machine Learning Research, California, Los Angeles, USA - 2009
Fecha de acceso: 19/03/2021
- 45 - **Confronting the Curse of Dimensionality**
Article of Open Data Science, USA - 2019
Fecha de acceso: 20/03/2021
- 46 - **Heurísticas para Data Augmentation en NLP**
RISTI, Revista Ibérica de Sistemas de Tecnología de Información, Portugal - 2019
Fecha de acceso: 22/03/2021
- 47 - **Curse of Dimensionality**
DeepAI, USA - 2019
Fecha de acceso: 22/03/2021
- 48 - **A Review of FS and FE Methods Applied on Microarray Data**
Zena Hira and Duncan Gillies
Department of Computing, Imperial College London, London, UK - 2015
Fecha de acceso: 22/03/2021
- 49 - **A Survey of feature selection and feature extraction techniques in ML**
Samina Khalid, Tehmina Khalid and Shamila Nasreen
Conference: Science and Information (SaI), London, UK - 2014
Fecha de acceso: 23/03/2021
- 50 - **Improving SVM Accuracy by Training on Auxiliary Data Sources**
Pengcheng Wu and Thomas Dietterich
School of EECS, Oregon State University, USA - 2004
Fecha de acceso: 05/04/2021
- 51 - **A Comparative Study of Methods for Transductive Transfer Learning**
Andrew Arnold, Ramesh Nallapati and William W. Cohen
Machine Learning Department, Carnegie Mellon University, Pittsburgh, USA - 2007
Fecha de acceso: 05/04/2021
- 52 - **Constructing Informative Priors using Transfer Learning**
Rajat Raina, Andrew Ng and Daphne Koller
Computer Science Department, Stanford University, USA - 2006
Fecha de acceso: 05/04/2021

53 - **Co-clustering based Classification for Out-of-domain Documents**

Wenyuan Dai, Gui-Rong Xue, Qiang Yang y Yong Yu
Shanghai Jiao Tong University, Shanghai, China - 2007
Fecha de acceso: 07/04/2021

54 - **Técnicas de regularización en regresión: implementación y aplicaciones**

Carrasco María
Departamento de estadística e investigación operativa Sevilla, España - 2016
Fecha de acceso: 04/05/2021