

# Modelo de Aprendizaje Continuo y Profundo para la Clasificación de Sentimientos Basados en Aspectos

## Investigación Doctoral

MSc. Dionis López Ramos<sup>1,2</sup>

<sup>1</sup>Centro de Investigación en Informática  
Universidad Central “Marta Abreu” de Las Villas

<sup>2</sup>Departamento de Ingeniería en Informática  
Universidad de Oriente

**Tutor:** Dr.C. Fernando José Artigas Fuentes

Febrero 2023

# Temas

- 1 Introducción
- 2 Modelo de Aprendizaje Continuo y Profundo para la extracción de aspectos
- 3 Modelo de Aprendizaje Continuo y Profundo para la clasificación de aspectos
- 4 Análisis de sentimientos basado en aspectos para la gestión de campañas de bien público
- 5 Consideraciones necesarias para extender los modelos propuestos al idioma español
- 6 Conclusiones

# Problema de investigación

El estudio de las opiniones o análisis de sentimientos es importante en diversas áreas de la sociedad<sup>1</sup>.



Turismo



Análisis automático de opiniones



Salud



Gobierno



<sup>1</sup>Bin Liu, (2020), "Sentiment analysis: Mining opinions, sentiments, and emotions": Cambridge university press.

# Problema de investigación

Los modelos de Aprendizaje Profundo en el Análisis de Sentimientos realizan la extracción y clasificación de opiniones **en un solo dominio del conocimiento** (p.ej., opiniones sobre hoteles).

- Al entrenar un modelo en diferentes dominios (e.d., opiniones de restaurantes, hoteles, efectos electrodomésticos, etc) su efectividad disminuye.
- Aparece el reto de reducir o evitar el **olvido catastrófico**.

## Aspect Based Sentiment Analysis (**ABSA**).

Subtarea del Análisis de Sentimientos orientada a la extracción y clasificación de información subjetiva a partir de contenidos generados por los usuarios. Por ejemplo:

- **La calidad** de *la ropa* es buena pero **las colas** para comprarlas son terribles.
- *El hotel* tiene **buen confort** pero **los elevadores** no son seguros.

# Análisis de Sentimiento Basado en Aspectos

En ABSA, dada una oración (secuencia de palabras)  $w^c = (w_1^c, \dots, w_n^c)$ :

- Una secuencia de aspectos está definida como  $w^t = (w_1^t, \dots, w_m^t)$  donde  $w^t$  es una sub-secuencia de  $w^c$ .
- El objetivo de esta tarea es predecir la polaridad del sentimiento  $y$  de la oración  $w^c$  hacia el aspecto objetivo  $w^t$ , donde  $y \in \{positivo, negativo, neutral\}$ .

## Objetivo de la investigación doctoral

Lograr el diseño e implementación de modelos de aprendizaje computacional **eficaces** y **eficientes** para la extracción y clasificación de aspectos, en opiniones de múltiples dominios en **idioma inglés, basadas en modelos de aprendizaje profundo y continuo.**

## Hipótesis

Un modelo computacional que combina una arquitectura de aprendizaje profundo centrada en mecanismos de atención, y un modelo de aprendizaje continuo que siga una estrategia de regularización en la subtarea de análisis de sentimientos basado en aspectos, logra una clasificación **eficaz** y **eficiente** de aspectos en documentos en idioma inglés provenientes de diversos dominios.

# Fundamentación de la investigación doctoral

Los trabajos que usan modelos de Aprendizaje Profundo han mostrado una alta eficacia en varias tareas del Procesamiento del Lenguaje Natural (p.ej., traducción, ABSA, clasificación de documentos)<sup>23</sup>:

- Las Redes Neuronales Convolucionales (*Convolutional Neural Network; CNN*).
- La Memoria a Corto Plazo (*Long Short Term Memory; LSTM*).
- las Unidades Recurrentes Cerradas (*Gated Recurrent Units; GRU*) y sus variantes Bidireccionales.
- Los Mecanismos de Atención.
- Los Grafos Convolucionales Neuronales.

---

<sup>2</sup>Dionis López, and Leticia Arco (2019), "Aprendizaje profundo para la extracción de aspectos en opiniones textuales":Revista Cubana de Ciencias Informáticas,105–145.

<sup>3</sup> "Deep Learning for aspect-based sentiment analysis: a comparative review",(2022), Komang Wahyu Trisna, and Huang Jin Jie : Applied Artificial Intelligence, 1–37.

# Aprendizaje Continuo

- Durante el proceso de aprendizaje, los seres humanos usan el conocimiento anterior al enfrentarse a una nueva tarea o reto.
- En los modelos de aprendizaje automático al aprender una tarea y enfrentarse a otra aparece el **olvido catastrófico**<sup>4</sup>.
- Esto se asocia a un problema más general en las redes neuronales llamado el dilema de la estabilidad-plasticidad<sup>5</sup>.

---

<sup>4</sup> "Catastrophic interference in connectionist networks: The sequential learning problem", (1989), Michael McCloskey and Neal J. Cohen : Psychology of learning and motivation,24, 109 – 165.

<sup>5</sup> "Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control", (2012), Stephen T. Grossberg :Springer Science and Business Media, 3(4), 70.

# Aprendizaje Continuo

Dada una secuencia de conjunto de datos (dominios de conocimiento)  $D_1, D_2, \dots, D_M$  uno a la vez, el objetivo es entrenar un modelo computacional  $f_M : x \rightarrow y$  con buenos resultados de calidad para todos los  $D_m; m \leq M$ ; después de haber sido entrenado secuencialmente en las  $M$  tareas, donde  $x \in \mathbb{R}^N$  es el espacio de entrada,  $N$  es la dimensión del vector (e.d., el tamaño máximo del vector de entrada en una red neuronal),  $y$  es el valor de inferencia de la probabilidad<sup>a</sup>.

---

<sup>a</sup> "Catastrophic interference in connectionist networks: The sequential learning problem", (1999), Robert M. French : Trends in cognitive sciences, 3(4), 128–135.

# Aprendizaje Continuo

Tiene como restricción no poder acceder a los datos de la tarea previa o a una cantidad limitada de información.

Si el modelo es re-entrenado con el conjunto de datos (dominio de conocimiento)  $D_M$ , este pudiera olvidar predecir el conjunto de datos  $D_m; m < M$  (**olvido catastrófico**)

# Componentes de un modelo de Aprendizaje Continuo

Administrador de dominios o tareas:

Dominio actual

$D_1, D_2, D_3, \dots, D_N, D_{N+1}, D_{N+2}, \dots, D_K$

Dominios aprendidos

Dominios por aprender



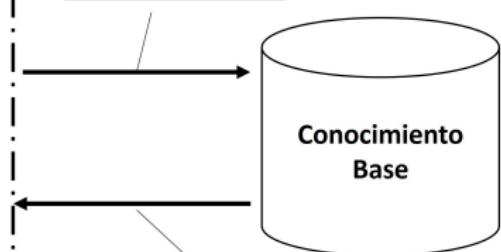
**Modelo Base:**

- CNN
- LSTM
- Mecanismo Atención
- BERT
- Otro



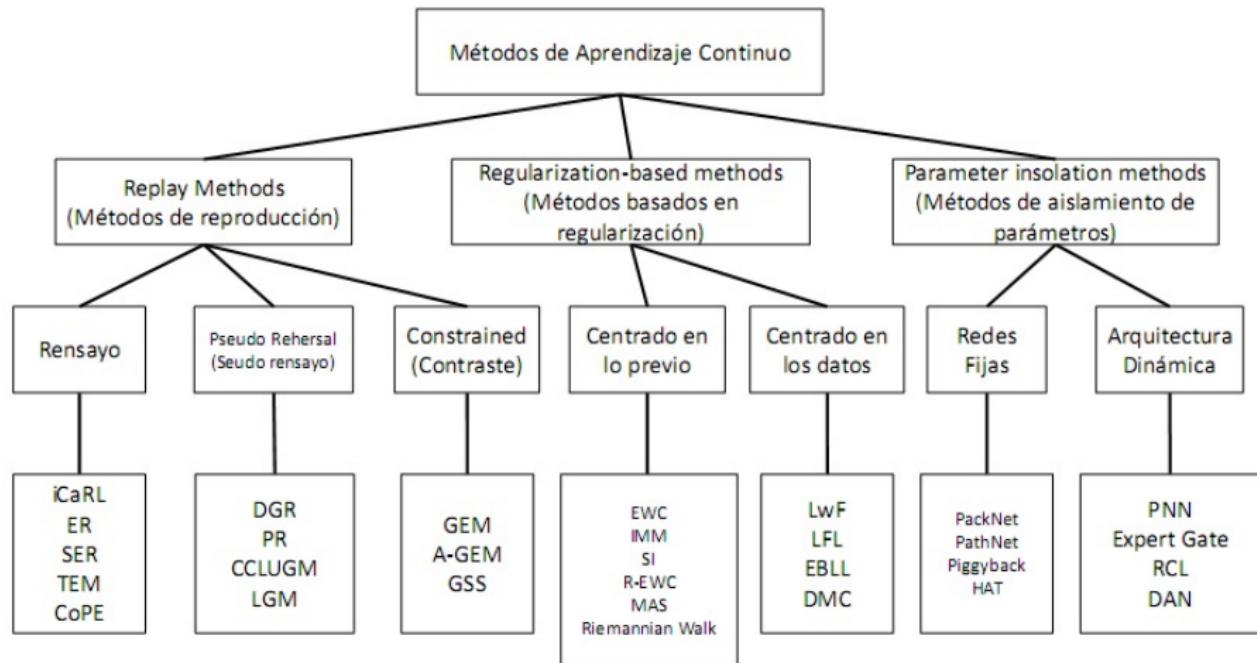
**Modelo de Aprendizaje Continuo**

Nuevo conocimiento



Conocimiento anterior

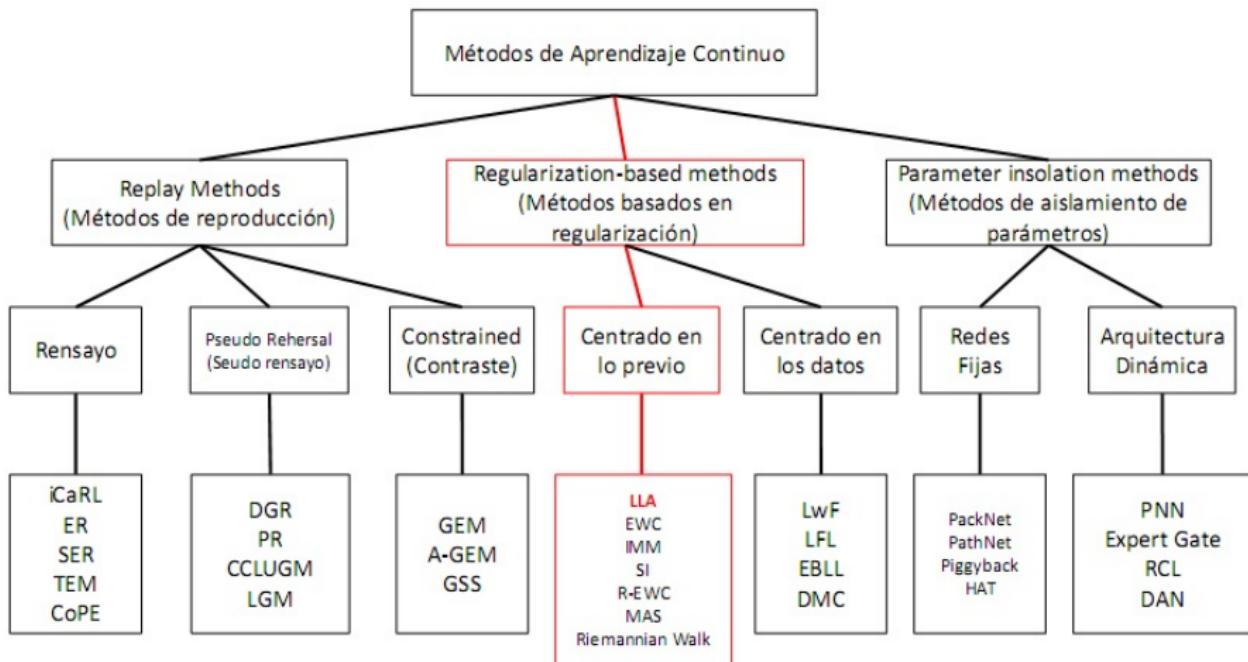
# Clasificación de los modelos de Aprendizaje Continuo



# Clasificación de los modelos de Aprendizaje Continuo

- **Aislamiento de parámetros (Parameter insolation)**: Dedica diferentes conjuntos de parámetros para cada tarea, puede crecer en nuevas ramas para cada nueva tarea.
- **Reproducción (Replay)**: Almacena ejemplos en formato crudo o genera seudo-ejemplos con un modelo generativo.
- **Regularización (Regularization)**: Estos métodos evitan almacenar entradas de datos en bruto, y reduce los requerimientos de memoria. Un término de regularización extra es introducido en la función de pérdida, consolidando el conocimiento previo cuando se aprenden nuevos datos.

# Clasificación de los modelos de Aprendizaje Continuo



- **Etapa 1: Pre-procesamiento del texto de una opinión:**

- ① Recibe la opinión textual original (determinar oraciones, entidades, etc).
- ② Retorna el vector de *tokens* para cada oración.
- ③ Crear la forma de representación computacional de los *tokens*.
- ④ Determinar cuáles son las palabras (*tokens*) que son aspectos.

## ¿Cómo reconocer los posibles aspectos?

- Etapa 2: Uso del modelo computacional para clasificar los aspectos en la oración.
- Etapa 3: Estimación de la polaridad global.
- Etapa 4: Visualización o almacenamiento de la información.

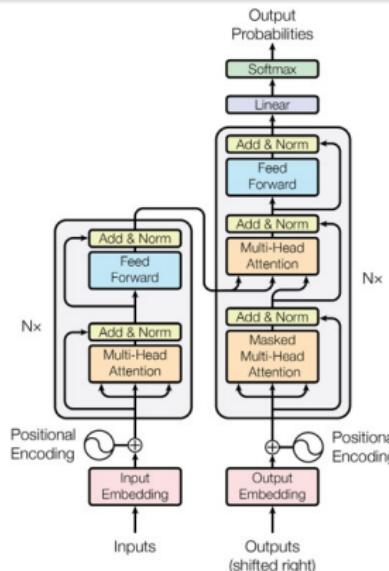
# Representación computacional de la información textual (Etapa 1)

- La Frecuencia Inversa por Documento (*Inverse Document Frequency*; TF-IDF).
- Vector de palabras embebidas (*Word Embeddings*).
  - Bolsa de Palabras Continuas (*Continuous Bag-of-Words*; CBOW).
  - Skip-Gram.
  - Vectores Globales (*Global Vectors*; Glove).
  - Modelos de Lenguaje Embebidos (*Embeddings from Language Models*; ELMo).
- **Modelo pre-entrenado de Representaciones del Codificador Bidireccional de Trasformadores (BERT; siglas en inglés).**

# Representación computacional de la información textual (Etapa 1)

## ¿Qué es BERT?

Modelo pre-entrenado de Representaciones del Codificador Bidireccional de Trasformadores (BERT; siglas en inglés).



# Modelo de Aprendizaje Continuo y Profundo para la extracción de aspectos (Etapa 1)

## Principales aportes

- Se combinan una arquitectura CNN con el modelo de aprendizaje continuo Aprendizaje sin olvido (Learning without forgetting; LWF)<sup>a</sup> en ABSA.
- La arquitectura de Aprendizaje Profundo está basada en la CNN propuesta por Poria et al en 2016<sup>b</sup>.
- Se crea una herramienta auxiliar al modelo computacional mediante el uso de reglas lingüísticas.

---

<sup>a</sup>Zhizhong Li, and Derek Hoiem (2018), "Learning without forgetting": Proceedings in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2935–2947

<sup>b</sup>Soujanya Poria, and Erik Cambria, and Alexander Gelbukh (2016), "Aspect extraction for opinion mining with a deep convolutional neural network": Knowledge-Based Systems, 108, 42–49.

# Diseño del Modelo (Etapa 1)

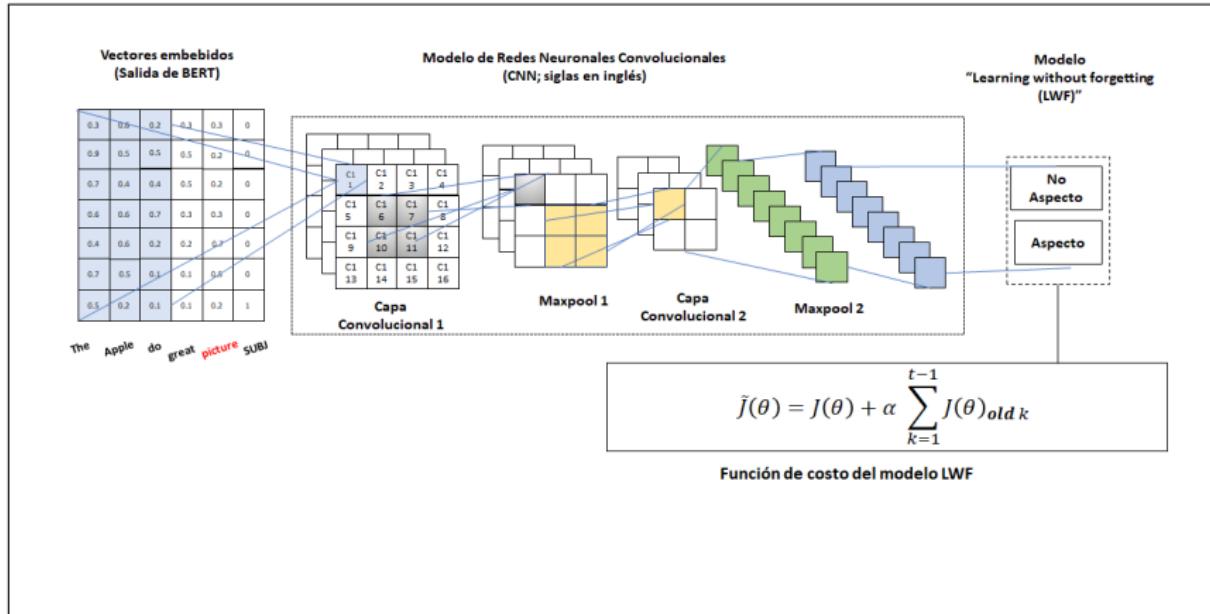


Figura: Esquema general del modelo para la extracción de aspectos Lwf-CNN.

# Algoritmo de Aprendizaje Continuo para la extracción de aspectos (Etapa 1)

---

## Algorithm 1

---

### Entrada:

$\Theta_s$  ▷ Parámetros compartidos del modelo CNN  
 $\Theta_p$  ▷ Parámetros de los dominios anteriores (capa de clasificación)  
 $c$  ▷ Valor del dominio actual  
 $X_c, Y_c$  ▷ Oraciones y aspectos etiquetados en el dominio  $c$  actual

### Salida:

$\Theta_s$  ▷ Modelo CNN  
 $\Theta_c$  ▷ La capa de salida del último dominio  
1:  $\Theta_c = \text{RandInit}()$  ▷ Nuevos parámetros inicializados aleatoriamente  
2: **for**  $x_i, y_i \leftarrow X_c, Y_c$  **do**  
3:    $y_p = \text{TrainCNN}(x_i, \Theta_s, \Theta_p)$  ▷ Salida de los dominios anteriores  
4:    $y_c = \text{TrainCNN}(x_i, \Theta_s, \Theta_c)$  ▷ Salida del dominio actual  
5:    $\text{Loss}'_i = \text{Loss}_i(y_c, y_i) + \alpha \sum_{k=1}^{c-1} \text{Loss}_{old}^k(y_p^k, y_k)$   
6: **end for**

---

# Evaluación de los resultados del entrenamiento del modelo (Etapa 1)

Para evaluar el modelo propuesto, se seleccionaron siete conjuntos de datos; con opiniones sobre dominios como:

- Reproductores de Mp3, DVD,
- Cámaras digitales y teléfonos inteligentes.
- Dos conjuntos de datos propuestos (e.d., un conjunto con opiniones acerca de restaurantes y otro sobre laptops) de SemEval 2014.
- Opiniones de hoteles del sitio TripAdvisor<sup>a</sup>.
- Las reglas lingüísticas fueron aplicadas a un conjunto de datos no etiquetado de opiniones en Amazon.

---

<sup>a</sup><http://times.cs.uiuc.edu/wang296/Data>

## Conjunto de reglas lingüísticas para ABSA

- Las “reglas lingüísticas” fueron definidas por los autores de [Poria, 2016] y propuestas en el propio trabajo.
- Definen cinco reglas que se apoyan en el lexicon SenticNet para extraer nuevos posibles aspectos de un conjunto de datos.
- Son un subconjunto de las reglas lingüísticas propuestas en [Qiu, 2011].

## Modelos del estado del arte comparados (Etapa 1)

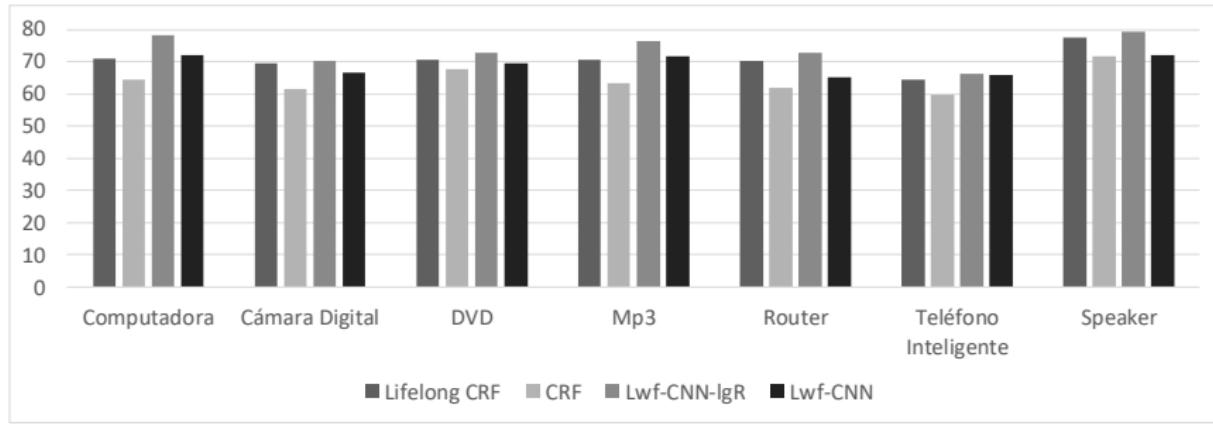
- **Lifelong CRF**: La estrategia presentada en Shu et al., 2016<sup>6</sup> que usa un modelo (*Conditional Random Field*; CRF) con una estrategia de Aprendizaje Continuo.
- **CRF**: Un modelo CRF evaluado en una estrategia de aprendizaje multitarea que, durante el entrenamiento, tiene en cuenta todos los dominios.
- **Lwf-CNN**: La propuesta presentada en este trabajo con el modelo CNN y la estrategia de Aprendizaje Profundo sin las reglas lingüísticas.
- **Lwf-CNN-IgR**: La propuesta presentada en este trabajo con el modelo CNN y la estrategia de Aprendizaje Profundo con las reglas lingüísticas.

---

<sup>6</sup>Lei Shu, and Bing Liu, and Hu Xu, and Annice Kim(2016) "Lifelong-rl: Lifelong relaxation labeling for separating entities and aspects in opinion targets":Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP),225–235.

# Resultados de las pruebas (Etapa 1)

Figura: Resultados de la medida F1-macro en las pruebas de dominio cruzado.



El modelo **Lwf-CNN-IgR** obtiene mejores valores de F1-macro al evaluarse en los diferentes conjuntos de datos.

# Resultados de las pruebas (Etapa 1)

**Cuadro:** Resultados en dominios de opiniones de restaurantes y hoteles  
(P:Exactitud, R:Exhaustividad y F1: F1-macro).

Entrenamiento	Pruebas	Lwf-CNN-IgR			Lwf-CNN		
		P	R	F1	P	R	F1
-Restaurante	Restaurante	74.8	53.7	62.5	69.8	46.8	56.0
-Hotel	Hotel	78.6	59.8	67.9	62.4	48.5	54.6

El modelo **Lwf-CNN-IgR** obtiene mejores valores de F1-macro, Exactitud y Exhaustividad al evaluarse en conjuntos de opiniones semánticamente distantes.

# Análisis de la eficiencia del modelo para la extracción de aspectos I

El *Lwf-CNN-IgR* combina el uso de un CNN como modelo base y un modelo para el AC de una capa completamente conectada a la salida del modelo CNN.

En un modelo CNN, la complejidad computacional por capas es de  $O(krd^2)$ . Donde  $r$  es el tamaño de la secuencia o vector de entrada a cada capa,  $d$  es la dimensión del vector y  $k$  es el tamaño del kernel de las convoluciones.

El costo computacional de una capa de neuronas completamente conectada es de  $O(n^2)$ . La capa de clasificación del modelo de AC para *Lwf-CNN-IgR* tiene una entrada de valor tres.

## Análisis de la eficiencia del modelo para la extracción de aspectos II

El costo computacional del aprendizaje del algoritmo para reglas lingüísticas se fue considerado como  $O(n)$ , se realiza sobre un único conjunto de entrenamiento.

La complejidad computacional total del modelo *Lwf-CNN-IgR* se define por:

$$O(mp \frac{n}{b} (krd^2)) \quad (1)$$

Para realizar los  $p$  ciclos y  $m > 2$  dominios a entrenar, los conjuntos de datos tienen una cantidad  $n$  de elementos máxima y una selección de  $b$  elementos en cada lote (batch) de entrenamiento.

# Análisis de la eficiencia del modelo para la extracción de aspectos III

El modelo CRF con el que es comparado el *Lwf-CNN-IgR*, tiene una complejidad computacional de  $O(cn^2)$ , donde  $n$  es la cantidad de ejemplos o instancias del conjunto de datos y  $c$  la cantidad de clases.

Para  $c \geq 2$  y  $b < n$  en el entrenamiento de un modelo de clasificación, el *Lwf-CNN-IgR* tiene una complejidad menor que el CRF:

$$O(mcn^2) \leq O(mp\frac{n}{b}) \quad (2)$$

# Resumen de la evaluación del modelo para la extracción de aspectos (Etapa 1)

- La propuesta presentada **Lwf-CNN-IgR** mejora los resultados del resto de los modelos comparados.
- El uso de las reglas lingüísticas permite incrementar el desempeño del modelo permitiendo ser una herramienta auxiliar al modelo de Aprendizaje Profundo.
- El modelo CNN propuesto en la investigación mejora los resultados del modelo CRF.
- El uso práctico de este modelo se orienta a ser parte del pre-procesamiento de oraciones en idioma inglés.
- Para determinar la posible polaridad (e.d., positivo, negativo o neutro) de los aspectos identificados por el modelo es necesario crear un clasificador de aspectos.

# Etapas para el uso de un modelo computacional en ABSA

- Etapa 1: Pre-procesamiento del texto de una opinión:
  - ① Recibe la opinión textual original (determinar oraciones, entidades, etc).
  - ② Retorna el vector de *tokens* para cada oración.
  - ③ Crear la forma de representación computacional de los *tokens*.
  - ④ Determinar cuáles son las palabras (*tokens*) que son aspectos.
- **Etapa 2: Uso del modelo computacional para clasificar los aspectos en la oración.**  
*¿Cómo determinar la polaridad o clasificación de los aspectos?*
- Etapa 3: Estimación de la polaridad global.
- Etapa 4: Visualización o almacenamiento de la información.

# Modelo de Aprendizaje Continuo y Profundo para la clasificación de aspectos (Etapa 2)

## Objetivo

Desarrollar un método que combine armónicamente las técnicas de procesamiento de lenguaje natural, el aprendizaje profundo y continuo; junto al análisis de sentimientos basado en aspectos

Para cumplir este objetivo se diseñó el modelo *LLA* siguiendo una estrategia de regularización de Aprendizaje Continuo.

## Principales aportes (Etapa 2)

- El modelo AR1 [Lomonaco, 2019] de regularización en el Aprendizaje Continuo, creado para la clasificación de imágenes, fue adaptado para la clasificación en ABSA.
- Se simplificó la forma de representación y entrada al pre-modelo BERT reduciendo la complejidad del modelo y mejorando su desempeño.
- Realiza el ajuste de los pesos de la red neuronal en BERT durante el proceso de aprendizaje.
- La preservación de pesos relevantes en la red neuronal se realiza durante el cálculo del Descenso del Gradiente.
- Los pesos de la capa del modelo de Aprendizaje Continuo son inicializados a 0 para el primer dominio (conjunto de datos).

## Diseño del Modelo Base (Etapa 2)

- El BERT Special (*BSp*) [Devlin, 2018] fue seleccionado como modelo base, por las ventajas en la representación de la información y su arquitectura.

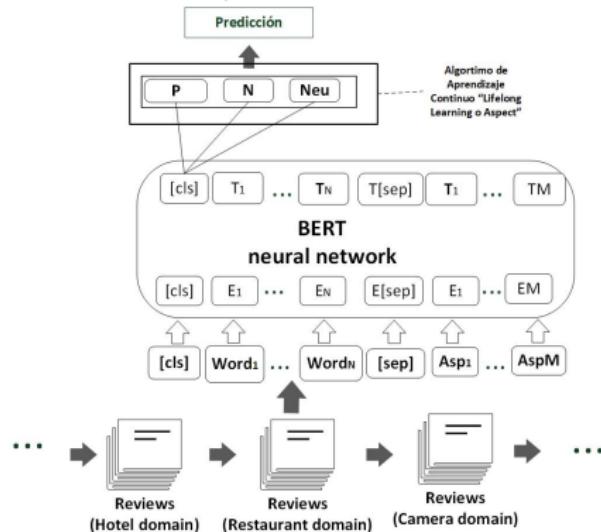
El modelo *BSp* construye una secuencia de entrada a la red neuronal de la forma:

$$\langle \text{CLS} \rangle + \text{tokens en la oración} + \langle \text{SEP} \rangle + [\text{asp}] + \langle \text{SEP} \rangle \quad (3)$$

- La arquitectura del modelo propuesto permite remplazar el modelo base por otra propuesta de Aprendizaje Profundo.

# Representación de la arquitectura del modelo para la clasificación de aspectos (Etapa 2)

**Figura:** Representación de la información de entrada y salida durante el proceso de aprendizaje (P: positivo, N: negativo, Neu: neutro).



## Función de pérdida del modelo (Etapa 2)

$$Loss'_t = Loss_t + \alpha \sum_k \Omega_k^t (\theta'_k - \theta_k)^2$$

Donde  $t$  representa el  $t$ -ésimo dominio,  $k$  son los parámetros de la arquitectura de red empleada,  $\alpha$  es un parámetro de fuerza que compensa las tareas antiguas (dominios) con las nuevas,  $\Omega_k^t$  representa la fuerza de regularización por parámetro y  $(\theta'_k - \theta_k)$  es el peso de referencia correspondiente a los parámetros al final del dominio anterior y el dominio actual.

# Apuntes sobre el modelo de clasificación de aspectos (Etapa 2)

- El *LLA* no extiende, a diferencia del modelo AR1, ninguna clase para cada nuevo dominio (e.d., se mantienen las clases positivo, negativo y neutral para cada aspecto).
- El principal objetivo en *LLA* es obtener los pesos ajustados para la capa de clasificación y ajustar los pesos de BERT con las instancias de cada dominio.
- Se reduce el olvido catastrófico mediante el uso del algoritmo Synaptic Intelligence (*SI*)<sup>7</sup>, para estimar la importancia de los pesos durante el cálculo del Descenso del Gradiente.

---

<sup>7</sup>Friedemann Zenke, and Ben Poole, Ben and Surya Ganguli (2017), "Continual learning through synaptic intelligence": arXiv preprint arXiv:1703.04200

# Evaluación de los resultados del entrenamiento del modelo (Etapa 2)

Cuadro: Conjuntos de datos (dominios de conocimiento) etiquetados.

Dominio	Oraciones	Aspectos	Entrenamiento	Prueba
Cámaras Digitales	597	237	477	120
Smartphones	546	302	436	110
Routers	701	307	877	176
Altavoces	687	440	549	138
Restaurantes	3841	4722	3041	800
Laptops	3845	2951	3045	800
Hoteles	4856	3810	3371	1485

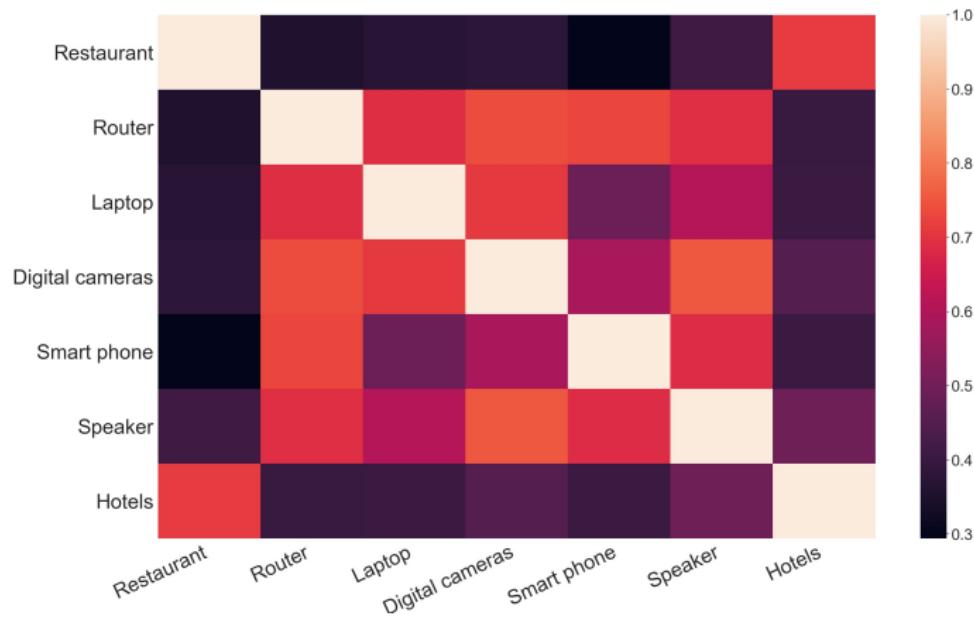
## Relación semántica de los conjuntos de datos (Etapa 2)

- Los conjuntos de datos utilizados (dominios de conocimiento) son considerados agrupamientos.
- Se calculó el vector centroide a partir de la media de los vectores de salida de BERT para las oraciones en cada dominio.
- Se calculó la semejanza coseno entre dos centroides (entre todos los dominios).

El coeficiente de Silhouette con respecto a la función de semejanza coseno fue de -0,017 (e.d., valores cercanos a cero indican solapamiento entre agrupamientos).

# Gráfica de la cercanía semántica entre dominios (Etapa 2).

Figura: Semejanza conseno entre el centroide de cada dominio.



## Medidas para evaluar el desempeño I (Etapa 2)

Teniendo en cuenta que los conjuntos de datos (dominios de conocimiento) seleccionados no están balanceados se emplearon:

### Exactitud (Accr)

$$\frac{TP + TN}{TP + TN + FP + FN}$$

$TP$  = Verdaderos Positivos,  $TN$  = Verdaderos Negativos

$FP$  = Falsos Positivos,  $FN$  = Falsos Negativos,

### F1-macro (F1)

$$F_1 = \frac{\text{Precisión} \times \text{Exhaustividad}}{\text{Precisión} + \text{Exhaustividad}}$$

## Medidas para evaluar el desempeño II (Etapa 2)

### Cohen-Kappa (Kappa)

$$\mathcal{K} = (\rho_o - \rho_e) / (1 - \rho_e)$$

$\rho_o$  = Probabilidad obtenida

$\rho_e$  = Valor real asignado

# Medida para la evaluación del olvido catastrófico (Etapa 2)

- Teniendo en cuenta el estado del arte de [Biesialska, 2020], la medida seleccionada fue propuesta en [Ke, 2020] y se consideró por:
  - ① Haber sido empleada en la evaluación de un modelo de Aprendizaje Continuo para el Análisis de Sentimiento.
  - ② Su definición tiene en cuenta la estimación de la calidad de la clasificación de los conjuntos aprendidos anteriormente.

## OvrcForgtt

El promedio del resultado de la clasificación final de los conjuntos de pruebas de las tareas anteriores a la última.

# Modelos del estado del arte comparados (Etapa 2)

## Aprendizaje Continuo

- Lifelong Learning Memory (*LLM*)
- Elastic Weight Consolidation (*EwC*)
- Architectural and Regularization 1 (*AR1*)

## Modelos Base

- Local Context Focus with BERT (*LC*)
- Attentional Encoder Network with BERT(*AE*)
- Attentional Encoder Network with LSTM (*AT*)

## Orden de entrenamiento de los conjuntos de datos de opiniones

- **REH:** Restaurantes, cámaras digitales, routers, speakers, laptops, Hoteles.
- **HER:** Hoteles, cámaras digitales, routers, speakers, laptops, restaurantes .
- **EHR:** Cámaras digitales, routers, speakers, laptops, hoteles, Restaurantes .
- **ERH:** Cámaras digitales, routers, speakers, laptops, restaurante, Hoteles .
- **RHE:** Restaurantes, hoteles, cámaras digitales, routers, speakers, laptops.
- **HRE:** Hoteles, restaurantes, cámaras digitales, routers, speakers, Laptops.

# Promedio de los resultados al emplear diferentes modelos base y el nuevo modelo de Aprendizaje Continuo (Etapa 2)

Model	$AE_{LLA}$	$AT_{LLA}$	$LC_{LLA}$	$BS_{PLLA}$
Accr	0.69	0.64	0.79	<b>0.80</b>
F1	0.49	0.38	0.66	<b>0.73</b>
Kappa	0.40	0.12	0.59	<b>0.62</b>
OvrcForgtt	0.49	0.38	0.66	<b>0.73</b>

# Promedio de los resultados al emplear diferentes modelos base y diferentes modelos de Aprendizaje Continuo (Etapa 2)

Model	LLM	$AT_{EwC}$	$AE_{AR1}$	$BSP_{EwC}$	$AT_{EwC}$	$LC_{AR1}$	$BSP_{LLA}$
Accr	0.39	0.68	0.57	0.75	0.61	0.64	<b>0.80</b>
F1	0.23	0.50	0.33	0.62	0.35	0.44	<b>0.73</b>
Kappa	0.03	0.42	0.16	0.53	0.11	0.32	<b>0.62</b>
OvrcForgtt	0.23	0.49	0.32	0.62	0.35	0.48	<b>0.73</b>

# Prueba de ablación del nuevo modelo

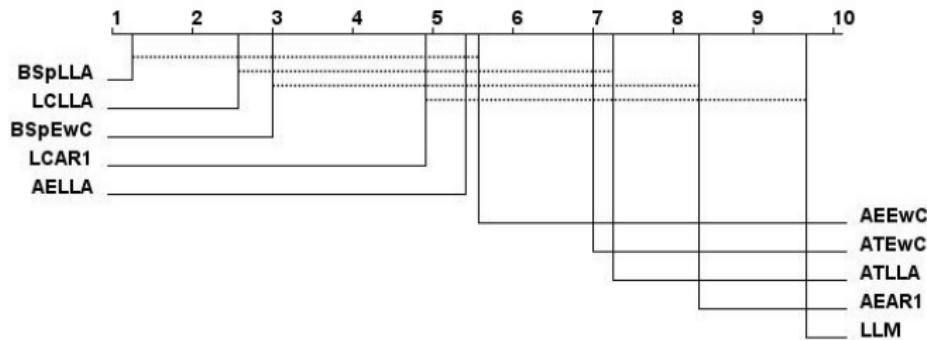
Modelo	$BSp$	$BSp_{LLA}$
Accr	0.64	<b>0.80</b>
F1	0.52	<b>0.73</b>
Kappa	0.39	<b>0.62</b>
OvrcForgtt	0.52	<b>0.73</b>

Los resultados muestran que el algoritmo *LLA* tiene un influencia importante en los resultados de clasificación y no puede ser eliminado porque provoca pérdida de eficacia.

## Prueba para evaluar la influencia del orden durante el entrenamiento del modelo (Etapa 2)

- Se hicieron las pruebas de Friedman y el método de Holm, para análisis post hoc con las medidas de desempeño F1.
- El objetivo fue identificar si existe influencia en el desempeño del modelo debido al orden del entrenamiento de los conjuntos de datos (dominios).

## Prueba de Holm para la medida de F1-macro (Etapa 2)



**Figura:** Evaluación de los modelos con la prueba de Holm con un nivel de especificidad de 0.05 para la medida de F1-macro

Para los valores de F1-macro en la prueba de Holm no se muestran grandes diferencias entre los modelos.

## Análisis de la influencia del orden (Etapa 2)

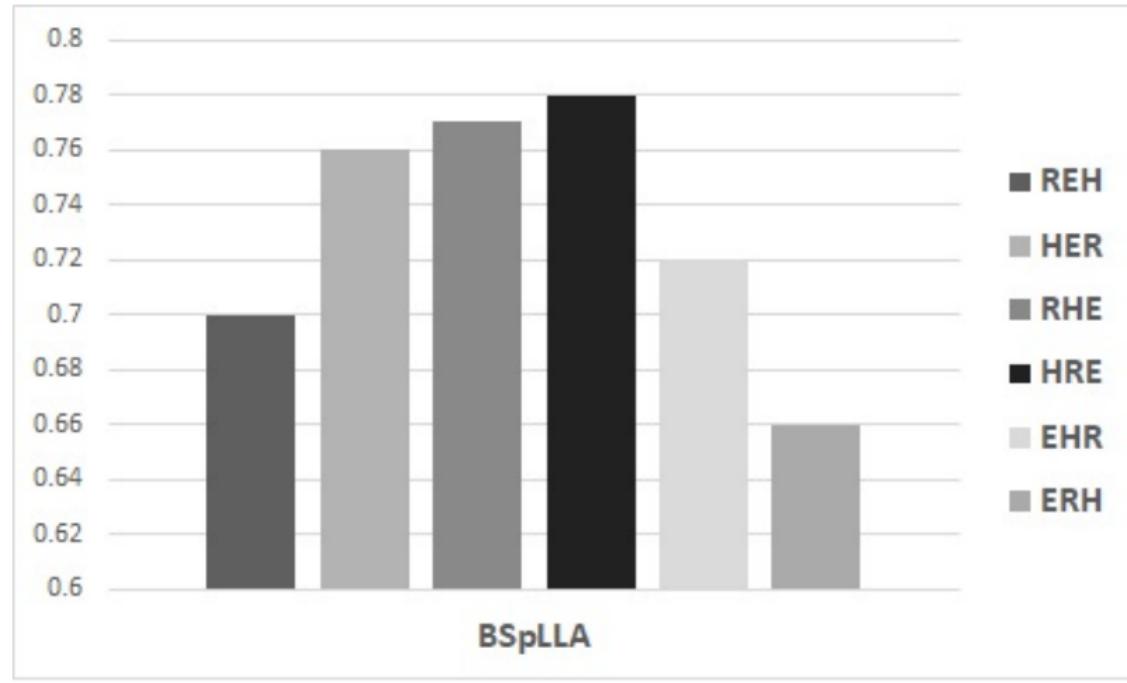


Figura: Resultados de  $BSp_{LLA}$  para los seis ordenamientos de los conjuntos de entrenamiento.

# Evaluación de la propuesta LLA y una propuesta reciente del estado del arte (Etapa 2)

- Se comparó la propuesta nombrado CLASSIC<sup>8</sup> que constituye una de las propuesta más reciente del estado del arte (e.d., Diciembre, 2021).
- En CLASSIC se propone un modelo que sigue la estrategia de Aprendizaje por Contraste, modificando la arquitectura de BERT en dos puntos (e.d., se agregan dos capas de red completamente conectadas).

---

<sup>8</sup> "Classic: Continual and contrastive learning of aspect sentiment classification tasks", (2021), Zixuan Ke, and Bing Liu, and Hu Xu, and Lei Shu : arXiv preprint arXiv:2112.02714,

## Valores de desempeño en CLASSIC declarados por los autores (Etapa 2)

Cuadro: Resultados de CLASSIC (según artículo) y los de *BSpLLA*.

Modelo	CLASSIC	<i>BSpLLA</i>
Accr	0.90	<b>0.80</b>
F1	0.85	<b>0.73</b>

# Análisis crítico del modelo CLASSIC I (Etapa 2)

- Se realizó la experimentación con 19 conjuntos de datos (e.d., los usados para entrenar  $BSpLLA$  son parte de los 19).
- Se seleccionaron cinco conjuntos de datos de forma aleatoria para estimar los resultados experimentales.
- No se especificó la cantidad de permutaciones entre conjuntos de datos que se usaron para estimar los resultados.
- Se realizaron ajustes diferentes de los hiperparámetros del modelo computacional según los conjuntos de datos a entrenar.

## Análisis crítico del modelo CLASSIC II (Etapa 2)

- Durante el análisis del código de la implementación de CLASSIC, se determinó que usaron una forma de entrada a la arquitectura BERT distinta a la que se usó en LLA.
- Para estimar el olvido catastrófico se usó la medida propuesta en Chaudhry et al., 2018<sup>9</sup>, que difiere a la usada en LLA.
- En CLASSIC, no se realizó el análisis de la cercanía semántica entre los conjuntos de datos.

---

<sup>9</sup> Arslan Chaudhry, and Puneet K. Dokania, and Thalaiyasingam Ajanthan, and Philip Torr (2018), "Riemannian walk for incremental learning: Understanding forgetting and intransigence": Proceedings of the European Conference on Computer Vision (ECCV), 532–547

# Definición de experimentos para evaluar CLASSIC y *BSpLLA* (Etapa 2)

A partir de estas diferencias se realizó un análisis comparativo de ambos modelos teniendo en cuenta los siguientes criterios:

- Comparación de ambos modelos (e.d., *CLASSIC* y *BSpLLA*) sobre los mismos conjuntos de datos de entrenamiento con un estudio de su cercanías semánticas.
- Uso de la misma medida para estimar el olvido catastrófico (usando la propuesta en *CLASSIC*).
- Uso del mismo conjunto de hiperparámetros para todos los conjuntos de datos como en *BSpLLA*.

# Resultados de los experimentos entre CLASSIC y $BSp_{LLA}$ (Etapa 2)

Cuadro: Resultados de la propuesta de CLASSIC y los de  $BSp_{LLA}$ .

Experimento	CLASSIC	$BSp_{LLA}$	Información
<i>Same-phd</i>	0.311	<b>0.316</b>	Igual conjuntos de datos.
<i>Same-parameters</i>	0.182	<b>0.316</b>	Igual hiperparámetros.
<i>Invert-input</i>	0.311	<b>0.316</b>	Invertir la forma de entrada.

En los experimentos se muestra que cuando se mantienen los mismos hiperparámetros (*Same-parameters*) se alcanza un mayor desempeño para  $BSp_{LLA}$ .

# Conclusión de los resultados de los experimentos entre CLASSIC y $BSp_{LLA}$ (Etapa 2)

- Los resultados experimentales muestran que  $BSp_{LLA}$  obtiene mejores resultados al reducir el olvido catastrófico.
- El mejor desempeño se asocia a que  $BSp_{LLA}$  evita grandes cambios en el valor de los pesos de la red neuronal durante el cálculo del Descenso del Gradiente.
- El ajuste de los pesos aislado en componentes de la red neuronal BERT (como se propone para CLASSIC) no es mejor que un modelo de Aprendizaje Continuo basado en regularización.

# Análisis de la eficiencia del modelo para la clasificación de aspectos I

- El modelo LLA (que utiliza BERT) es entrenado para  $m$  conjuntos de datos o dominios de entrenamiento.
- La complejidad por capas de BERT es  $O(r^2d + rd^2)$ .
- La arquitectura de BERT empleada tiene 12 capas.
- Para una complejidad de  $12 * O(r^2d + rd^2)$ , esta se reduce a  $O(r^2d + rd^2)$ .

El modelo LLA tiene un costo computacional de:

$$O(mp \frac{n}{b} (r^2d + rd^2)) \quad (4)$$

## Análisis de la eficiencia del modelo para la clasificación de aspectos II

En las experimentaciones realizadas se compara el LLA con tres modelos de AC (AR1, EWC, CLASSIC) usando el mismo modelo base (e.d., BERT).

LLA tiene una complejidad computacional mayor que AR1 por el ajuste de los pesos de la capa del modelo de AC y el uso de BERT.

El EWC es más costoso que el LLA porque emplea en el cálculo del descenso del gradiente la matriz de Fisher. Su complejidad aproximada es  $O(r^2 \log r)$ .

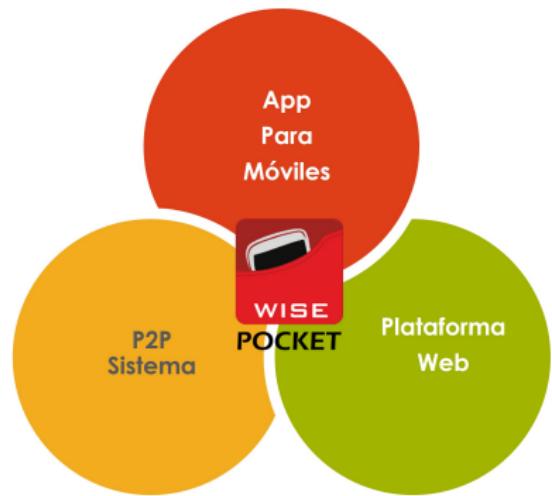
El modelo CLASSIC, tiene menor complejidad computacional que LLA en dos capas de red neuronal de la arquitectura de su modelo.

- Etapa 1: Pre-procesamiento del texto de una opinión:
  - ① Recibe la opinión textual original (determinar oraciones, entidades, etc).
  - ② Retorna el vector de *tokens* para cada oración.
  - ③ Crear la forma de representación computacional de los *tokens*.
  - ④ Determinar cuáles son las palabras (*tokens*) que son aspectos.
- Etapa 2: Uso del modelo computacional para clasificar los aspectos en la oración.
- **Etapa 3: Estimación de la polaridad global.**
- **Etapa 4: Visualización o almacenamiento de la información.**

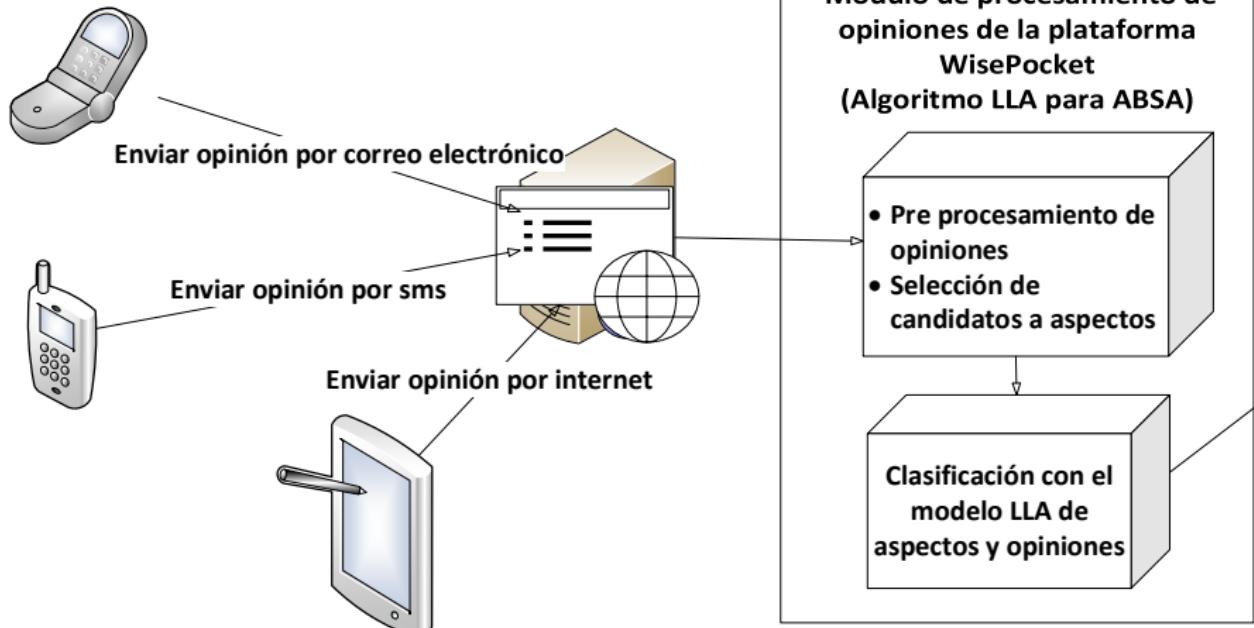
# La gestión de la información en las campañas de bien público (Etapa 3 y 4)

- Las campañas sociales o de bien público son una herramienta importante para promover un cambio positivo en las actitudes sociales.
- Las encuestas de opinión permite la retroalimentación de su efectividad.
- El Análisis de Sentimientos de las encuestas de opinión permite mejorar los resultados de una campaña.

# WisePocket: Plataforma digital para campañas de bien público (Etapa 3 y 4)



# WisePocket: Plataforma digital para campañas de bien público (Etapa 3 y 4)



# Caso de uso: Campaña de bien público para la divulgación del XIII Seminario Nacional sobre Estudios Canadienses (Etapa 3 y 4)

Los modelos computacionales propuestos en esta investigación se utilizaron en el XIII Seminario Nacional sobre Estudios Canadienses el 21 de abril de 2022, como parte del módulo de Análisis de Sentimientos de la plataforma WisePocket.

En el procesamiento de opiniones en idioma inglés se lograron procesar y clasificar los aspectos de unas 50 opiniones. Detectando 120 aspectos (50 positivos, 30 negativos y 40 neutrales).

# Caso de uso: Campaña de bien público para la divulgación del XIII Seminario Nacional sobre Estudios Canadienses (Etapa 3 y 4)

- En 7 opiniones el algoritmo dio una clasificación errónea (p.ej. “The conference was perfect but the coffee break was incredible” donde *perfect* fue clasificada como **positiva** e *incredible* **negativa**).
- Se realizaron 68 envíos de opiniones y en 11 de los casos el algoritmo de extracción no detectó los posibles aspectos (desechados como posibles opiniones).
- 7 opiniones fueron enviadas en español como si fueran opiniones en inglés (error humano).
- De los 11 casos el modelo de extracción no detectó 7 posibles opiniones y el resto de los casos fueron errores ortográficos enviados por los usuarios.

# Consideraciones necesarias para extender los modelos propuestos al idioma español

- Evaluar el uso de pre-modelos para el español basados en la arquitectura BERT (huggingface.co):
  - ① BETO: Spanish BERT.
  - ② RoBERTa base entrenado con datos de la Biblioteca Nacional de España (BNE).
  - ③ XLM-RoBERTa (base-sized model).
- Estudiar el uso de reglas lingüísticas y otros recursos para minar conjuntos de datos como las reglas propuestas en Qiu et al., (2009)<sup>a</sup>.

---

<sup>a</sup> "Expanding Domain Sentiment Lexicon through Double Propagation", International Joint Conference on Artificial Intelligence, (2009):Guang Qiu , Bing Liu, Jiajun Bu, and Chun Chen.

# Consideraciones necesarias para extender los modelos propuestos al idioma español II

- Buscar o construir conjuntos de datos de opiniones orientadas a aspectos para diferentes dominios en idioma español.
- Evaluar el uso de los modelos (Learning without Forgetting with Linguistic Rules; Lwf-CNN-IgR) y (Lifelong Learning of Aspects; LLA) para el idioma español en ABSA.
- Uso de la herramienta **spaCy** y sus módulos para el idioma español (etiquetador morfológico, reconocedor de entidades, divisor de oraciones, etc).
- Tener una institución interesada en la aplicación de los resultados (e.d., **utilidad para Cuba**). En el transcurso de esta investigación doctoral se propuso extender al español los resultados a:
  - ① Periódico Granma (Sección “Cartas al Director”).
  - ② Delegación Provincial de Turismo de Santiago de Cuba.

# Otros escenarios en Cuba para el uso de los resultados de la investigación

## El turismo

La herramienta propuesta en esta investigación, permitiría lograr el análisis de la información asociada a las opiniones en idioma inglés que los turistas emiten sobre las instalaciones hoteleras y otros lugares de esparcimiento.

## Análisis de noticias

En instituciones de defensa del país o donde su objeto social se relaciona con el procesamiento y análisis de información de fuentes internacionales en idioma inglés (p.ej., Centro de Prensa Internacional).

# Conclusiones de la Investigación Doctoral

Se diseñaron dos modelos computacionales:

- ① El *Learning without Forgetting with Linguistic Rules* (Lwf-CNN-IgR) para la extracción de aspectos en un flujo de opiniones, de diferentes dominios, en idioma inglés.
- ② El *Lifelong Learning of Aspects* (LLA) para la clasificación de aspectos en un flujo de información de iguales características al anterior.

# Conclusiones de la Investigación Doctoral II

Se cumplió el objetivo general planteado, ya que:

- Se usó la estrategia de regularización para la combinación de modelos de Aprendizaje Profundo y Continuo reduciendo el olvido catastrófico, el gasto de memoria y el tiempo de entrenamiento.
- Se creó un modelo para la extracción de aspectos que permite mejorar la eficacia y calidad del preprocesamiento de la información.

# Conclusiones de la Investigación Doctoral III

- La regularización de los cambios en los pesos de la arquitectura de red, durante el proceso de Aprendizaje Continuo tiene una influencia positiva en la reducción del olvido catastrófico.
- Se obtuvieron valores de eficacia para la clasificación de aspectos ( $F1\text{-macro} = 0.73$ ) y la reducción del olvido catastrófico (0.316) que superan a otras propuestas recientes del estado del arte.
- El uso del modelo pre-entrenado BERT para la representación del conocimiento de los modelos de Aprendizaje Profundo y Continuo propuestos permitió obtener altos valores de calidad en la extracción y clasificación de aspectos.

# Recomendaciones y trabajo futuro

Derivadas del estudio realizado, así como de las conclusiones generales emanadas del mismo, se recomienda:

- ① Estudiar la posibilidad de crear una medida de desempeño común para la estimación del olvido catastrófico.
- ② Estudiar la posibilidad de combinar los métodos de regularización del cambio de los pesos en el proceso de aprendizaje de modelos de Aprendizaje Continuo con propuestas de adaptación de componentes de modelos pre-entrenados como BERT.
- ③ Extender el estudio a propuestas de modelos para el idioma español.
- ④ Estudiar el comportamiento del modelo de clasificación de aspectos con el apoyo de reglas lingüísticas para analizar el desempeño de la calidad en la clasificación.

# ChatGPT 2022 (consulta enero 2023): Sugerencias a seguir para esta investigación

- ¿Cómo se puede diseñar un algoritmo de Aprendizaje Profundo y Continuo para el Análisis de Sentimientos?
  - ① Recopilar y preprocesar un gran conjunto de datos de texto etiquetados con sentimiento (positivo, negativo, neutral).
  - ② Elejir una arquitectura de modelo de aprendizaje profundo adecuada, como una LSTM o una red de transformadores (p.ej., BERT).
  - ③ Entrenar el modelo en el conjunto de datos utilizando un algoritmo de optimización adecuado, como Adam o RMSprop.
  - ④ Actualizar continuamente el modelo utilizando un método de aprendizaje continuo, como Elastic Weight Consolidation (EWC) o Synaptic Intelligence (SI).
  - ⑤ Finalmente, evaluar el rendimiento del modelo utilizando métricas como la precisión, la puntuación F1 y la matriz de confusión.

# Muchas gracias...

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 1

Profundice en los siguientes elementos que sustentan la necesidad de desarrollar la investigación:

- a. ¿Qué elementos se tuvieron en cuenta para abordar el problema de análisis de sentimiento basado en aspecto respecto a un enfoque tradicional?.
- b. En diferentes momentos de la memoria escrita se hace referencia a la introducción de nuevos métodos o modelos a partir de la combinación de métodos del aprendizaje profundo y por continuación (continuos). Argumente hacia dónde va dirigido su contribución (Proponer nuevos métodos o modelos).

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 1: (Continuación)

- e. ¿Qué significa para la presente investigación que el problema sea un ABSA de múltiples dominios (Opiniones de temas libres o un entorno de múltiples dominios controlados como el escenario de experimentación que se propone)?.
- e. Ilustre posibles aplicaciones de su propuesto en el contexto cubano argumentando una posible solución para idioma español.
  - ! Profundice en la decisión de no haber considerado en la presente investigación una propuesta en idioma español siendo este un campo de mayor necesidad de desarrollo teórico.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 1: (Continuación)

- d. La competición REST-Mex 2022 puso a disposición de la comunidad científica un conjunto de datos de opiniones de tres dominios asociados al turismo, con textos en idioma español. Pudiera emplearse este conjunto de datos con uno de los pre-modelos propuestos en el capítulo 3 (p. ej., ROBERTA o xmlBERT) como parte de sus propuestas. **(No respondida)**

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Respuesta Pregunta 1 a.

Profundice en los siguientes elementos que sustentan la necesidad de desarrollar la investigación:

- a. ¿Qué elementos se tuvieron en cuenta para abordar el problema de análisis de sentimiento basado en aspecto respecto a un enfoque tradicional?.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 a.)

## Respuesta Pregunta 1 a. (Oración de ejemplo)

**El servicio de ventas** es el mejor de la ciudad, **los dependientes** son muy amables y los precios aceptables; aunque **el local** no tiene confort y ventilación.

## Respuesta Pregunta 1 a.

- **El servicio de ventas:**

- ① el mejor de la ciudad, (**positivo**).
- ② los precios aceptables (**positivo o neutral**).

- **los dependientes:**

- ① son muy amables (**positivo**)

- **el local:**

- ① no tiene confort (**negativo**).
- ② ventilación (**negativo**).

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 a.)

## Respuesta Pregunta 1 a.

El **Análisis de Sentimientos** es una tarea del Procesamiento del Lenguaje Natural [Zhang, Dic. 2022]

### Subtareas:

- Nivel de documento.
- Nivel de oraciones o partes del texto.
- **Análisis de Sentimiento basado en Aspectos (ABSA; siglas en inglés)** [López, 2019, Zhang, Dic. 2022]
  - ① Extracción del objeto de la opinión (Opinion Target Expression; OTE)
  - ② Detección de la categoría del aspecto (Aspect Category; AC)
  - ③ Polaridad del sentimiento (Sentiment Polarity; SP)

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

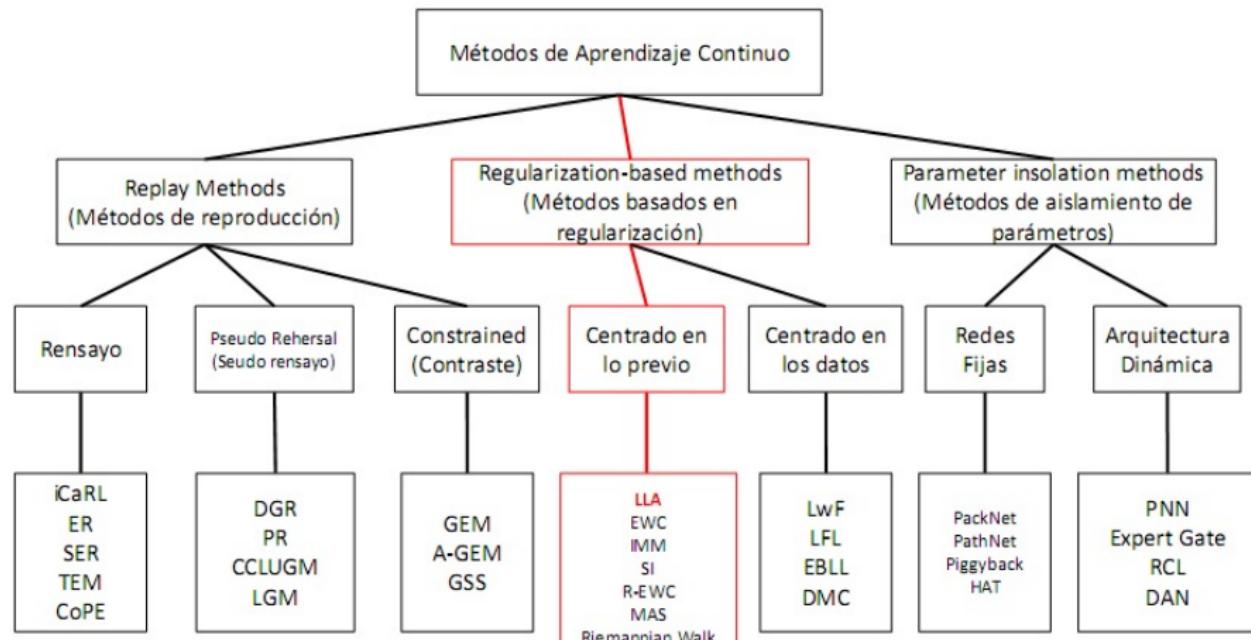
## Respuesta Pregunta 1 b.

Profundice en los siguientes elementos que sustentan la necesidad de desarrollar la investigación:

- b. En diferentes momentos de la memoria escrita se hace referencia a la introducción de nuevos métodos o modelos a partir de la combinación de métodos del aprendizaje profundo y continuo. Argumente hacia dónde va dirigido su contribución (Proponer nuevos métodos o modelos)

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 b.)

## Clasificación de los modelos de Aprendizaje Continuo Pregunta 1 b.



# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 b.)

## Respuesta Pregunta 1 b.

Esta investigación doctoral mediante la aplicación del **método basado en regularización** propone la combinación de un modelo de **aprendizaje profundo (AP)** y otro de **aprendizaje continuo (AC)** para resolver dos sub tareas del ABSA.

- Extracción de aspectos
- Clasificación de la polaridad de aspectos

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 c.)

## Respuesta Pregunta 1 c.

Profundice en los siguientes elementos que sustentan la necesidad de desarrollar la investigación:

- ¿Qué significa para la presente investigación que el problema sea un ABSA de múltiples dominios (Opiniones de temas libres o un entorno de múltiples dominios controlados como el escenario de experimentación que se propone)?.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 c.)

## Respuesta Pregunta 1 c.

Esta investigación se centra en los modelos de Aprendizaje Continuo en un entorno:

- **Aprendizaje Continuo de Dominios (DIL; siglas en inglés)** : Todas las tareas (p.ej., clasificación de opiniones de restaurantes, hoteles) comparten las mismas clases (p.ej., positivo, negativo, y neutral en el Análisis de Sentimiento).

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 c.)

## Componentes de un modelo de Aprendizaje Continuo en ABSA

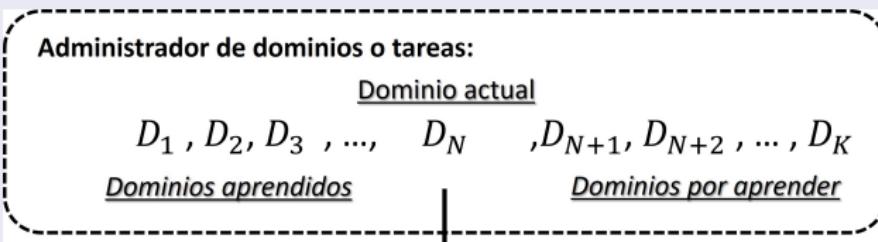
Administrador de dominios o tareas:

Dominio actual

$D_1, D_2, D_3, \dots, D_N, D_{N+1}, D_{N+2}, \dots, D_K$

Dominios aprendidos

Dominios por aprender



**Modelo Base:**

- CNN
- LSTM
- Mecanismo Atención
- BERT
- Otro

Nuevo conocimiento



Conocimiento Base

Conocimiento anterior

Modelo de Aprendizaje Continuo

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 c.)

## Respuesta Pregunta 1 c.

- Para el entrenamiento se agrupan las opiniones por dominios (p.ej., opiniones de restaurantes, efectos electrodomésticos, etc).
- El modelo propuesto aprende de manera secuencial de cada dominio o conjunto de datos de opiniones.
- Es importante que no se degrade la calidad de la predicción de (extracción/clasificación) opiniones de dominios anteriores.
- En un sistema de producción se podrá predecir opiniones de temas similares al de los dominios de entrenamiento.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 d.)

## Respuesta Pregunta 1 d.

Profundice en los siguientes elementos que sustentan la necesidad de desarrollar la investigación:

- d. Ilustre posibles aplicaciones de su propuesto en el contexto cubano argumentando una posible solución para idioma español. Profundice en la decisión de no haber considerado en la presente investigación una propuesta en idioma español siendo este un campo de mayor necesidad de desarrollo teórico.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 d.)

## Respuesta Pregunta 1 d.

Profundice en los siguientes elementos que sustentan la necesidad de desarrollar la investigación:

- d. Ilustre posibles aplicaciones de su propuesto en el contexto cubano argumentando una posible solución para idioma español.
  - 1. Profundice en la decisión de no haber considerado en la presente investigación una propuesta en idioma español siendo este un campo de mayor necesidad de desarrollo teórico.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 d.)

## Respuesta Pregunta 1 d.

Es posible crear un modelo experimental en español y su posterior uso en producción para el análisis de opiniones de la población, turistas y público en escenarios como:

- ① Periódico Granma (Sección “Cartas al Director”).
- ② Ministerio de Turismo (Delegaciones provinciales en inglés y español).
- ③ Plataformas del “Portal del ciudadano” en provincias de Cuba.
- ④ Órganos de análisis de información del Ministerio del Interior.
- ⑤ Asistentes virtuales de oficinas de atención a la población.
- ⑥ Otros.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 d.)

## Respuesta Pregunta 1 d.

Profundice en los siguientes elementos que sustentan la necesidad de desarrollar la investigación:

- ① Ilustre posibles aplicaciones de su propuesto en el contexto cubano argumentando una posible solución para idioma español.
- ② Profundice en la decisión de no haber considerado en la presente investigación una propuesta en idioma español siendo este un campo de mayor necesidad de desarrollo teórico.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 d.)

## Respuesta Pregunta 1 d.

Algunos elementos son:

- ① BETO[Cañete, 2020] es el primer pre-modelo basado en BERT para idioma español.
- ② Han surgido varios pre-modelos, para el español, basados en BERT o con arquitectura de Transformers.
- ③ No se cuenta con muchos conjuntos de datos con las características de la tarea ABSA para el español:
  - Texto de la opinión.
  - Aspecto (palabra en el texto).
  - Polaridad del aspecto.
- ④ Varias de las instituciones (Santiago de Cuba) consultadas para el uso de una solución en idioma español no mostraron interés en el uso de nuestra propuesta.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 e.)

## Respuesta Pregunta 1 e.

Profundice en los siguientes elementos que sustentan la necesidad de desarrollar la investigación:

- e. La competición REST-Mex 2022 puso a disposición de la comunidad científica un conjunto de datos de opiniones de tres dominios asociados al turismo, con textos en idioma español. Pudiera emplearse este conjunto de datos con uno de los pre-modelos propuestos en el capítulo 3 (p. ej., ROBERTA o xmlBERT) como parte de sus propuestas.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (1 e.)

## Respuesta Pregunta 1 e.

- El conjunto de datos (REST-Mex 2022) no está orientado a aspectos.
- En el texto de las opiniones no se especifica cuáles serían los posibles aspectos o características en las que se habla sobre una entidad nombrada y la polaridad que tiene este aspecto.
- “A priori” no podremos obtener buenos resultados por el inconveniente de no contar con un conjunto de datos orientado a aspectos (e.d., de análisis de sentimiento pero no orientado a aspectos).
- El uso de un pre-modelo como ROBERTA o xmBERT en el que su entrenamiento tuvo menor conjuntos de entrenamiento del usado para BERT.
- Forzar la polaridad propuesta por los etiquetadores no es recomendable (e.d., para REST-Mex 2022 se proponen 5 niveles).

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 2

A partir de la definición formal dada sobre AC en la página 5 explique, detalladamente y sobre la base una formalización matemática en que consiste el AC. Note que se han omitido aspectos relevantes en la formalización como **los espacios de partida y llegada**, los modelos aprendidos  $f_T$  parecieran aprenderse independientes o incluso se considera o **interpreta como un único modelo** el de la posición  $T$  entre otros.:

- a. Que sucedería con el olvido catastrófico y con la propuesta de solución si se consideraran no solo el dominio anterior  $t - 1$  sino más de un dominio por ejemplo  $t - 2$ . Como se formalizaría el problema de AC en caso de que el aspirante considere que una variante así tiene sentido.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (2 a.)

## Respuesta a Pregunta 2 a. [Ke, 2022]

En el Aprendizaje Continuo (AC) se aprenden una secuencia de tareas  $1, \dots, T$  incrementalmente. Cada tarea tiene un conjunto de entrenamiento  $D_\tau$ . Se denota cada ejemplo en  $D_\tau$  como  $(X_\tau, Y_\tau)$ , donde  $X_\tau$  es el conjunto de ejemplos de la tarea  $\tau$  y  $Y_\tau$  es la etiqueta correspondiente (para un esquema no supervisado no existiría  $Y_\tau$ ).

El objetivo del AC es minimizar la pérdida esperada de todas las tareas o dominios vistos limitando o no el acceso a las tareas anteriores  $\tau < T$ . Donde  $T$  es la cantidad de tareas de forma que:

$$\sum_{\tau=1}^T \mathbb{E}_{(X_\tau, Y_\tau)} [\mathcal{L}(f_\tau(X_\tau; \theta), Y_\tau)] \quad (5)$$

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (2 a.)

## Respuesta a Pregunta 2 a (Continuación)

Donde  $f_\tau$  representa la red para la tarea  $\tau$  y  $\theta$  los parámetros de la red/pesos. En la ecuación anterior es fácilmente aproximar para la tarea actual  $T$  porque los datos de la tarea actual están disponibles:

$$\frac{1}{n_T} \sum_{i=1}^{n_T} \mathcal{L}((f(x_\tau^i; \theta), y_\tau^i)) \quad (6)$$

Sin embargo, los datos de entrenamiento no están disponibles para las tareas anteriores de manera que la pérdida no puede ser aproximada usando.

El principal reto en AC es como determinar los parámetros  $\theta$  que optimicen la ecuación 5

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Respuesta a la Pregunta 2 a)

- a. Que sucedería con el olvido catastrófico y con la propuesta de solución si se consideraran no solo el dominio anterior  $t - 1$  sino más de un dominio por ejemplo  $t - 2$ . Como se formalizaría el problema de AC en caso de que el aspirante considere que una variante así tiene sentido.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)(2 a.)

## Respuesta a la Pregunta 2 a.

- Un modelo de aprendizaje continuo persigue el objetivo de no perder el conocimiento anterior, aun cuando aparezcan nuevos dominios en el aprendizaje incremental [Chen, 2018, Biesialska, 2020].
- Este proceso en un instante  $t$  incluye la posibilidad de que el modelo entrenado no pierda calidad en su desempeño (las medidas de calidad obtengan buenos valores) al ser evaluado en conjuntos de pruebas asociados a los dominios  $t-1$  y  $t-2$ .
- La medida de estimación del olvido catastrófico empleada en esta investigación relaciona el valor de la predicción del modelo entrenado en el dominio actual (instante  $t$ ), con su desempeño en los dominios anteriores ( $t-1, t-2, \dots, 1$ ).

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 3

Que elementos del campo de acción enunciado en la investigación son transformados en la presente investigación para arribar a resultados teóricos novedosos:

- a. Profundice de manera explícita en los elementos teóricos que sostienen la novedad de la investigación más allá de la combinación de dos métodos computacionales.
- b. Exponga mediante un experimento sencillo y reproducible las ventajas que aporta introducir el AC respecto a no tenerlo y solo usar una arquitectura de AP (**Implementación**).

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (3 a.)

## Respuesta 3 a

- a. Profundice de manera explícita en los elementos teóricos que sostienen la novedad de la investigación más allá de la combinación de dos métodos computacionales.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (3 a.)

## Respuesta 3 a

- a. Estado del arte actualizado sobre el tema son propuestos en [Biesialska, 2020, Ke, 2022]
- b. Hasta el conocimiento de esta investigación existen cuatro trabajos asociados al tema (ABSA y Aprendizaje Continuo) y se concentran en las publicaciones de los autores Zixuan Ke, **Bing Liu**, Chengyu Wang.
- c. Hasta el conocimiento de esta investigación no tenemos referencia de modelos que hayan empleado el algoritmo “Inteligencia Sináptica” y el modelo BERT.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (3 a.)

## Respuesta 3 a

- ④ El uso del método de regularización de AC y la compensación de los pesos en la arquitectura de red neuronal y no solo en su capa superior, para la creación de un modelo de AC para ABSA.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Respuesta 3 b

- a. Exponga mediante un experimento sencillo y reproducible las ventajas que aporta introducir el AC respecto a no tenerlo y solo usar una arquitectura de AP.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Respuesta 3 b

A pesar de los problemas presentados para ejecutar la implementación asociada a la respuesta a esta pregunta:

- Orientación de finalizar los trabajos en el cluster de HPC de la Universidad Jaume I de Castellón (Comunidad Valenciana, España).
- Interrupciones eléctricas inesperadas por más de tres horas en diferentes partes del país.
- 9 días para lograr estos resultados.
- Varias tareas de implementación asociadas a la otras preguntas de la oposición.

# Prueba de ablación del nuevo modelo

Modelo	$BSp$	$BSp_{LLA}$
Accr	0.64	<b>0.80</b>
F1	0.52	<b>0.73</b>
Kappa	0.39	<b>0.62</b>
OvrcForgtt	0.52	<b>0.73</b>

Los resultados muestran que el algoritmo *LLA* tiene un influencia importante en los resultados de clasificación y no puede ser eliminado porque provoca pérdida de eficacia.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 4

En la memoria escrita se afirma: "Para lograr que los modelos computacionales de AP puedan predecir la información de flujos de opiniones de diferentes dominios (e.d., opiniones sobre hoteles, restaurantes, elecciones presidenciales), es necesario crear modelos de AC [36]" :

- a. En este sentido podría entenderse que el aprendizaje continuo (AC) es la alternativa que el AP no es capaz de cubrir para abordar problemas de esta naturaleza. Exponga elementos acerca de este enfoque, teniendo en cuenta otras variantes del AP como pueden ser los modelos de atención con estructuras recurrentes como seq2seq u otras arquitecturas evaluadas en escenario de múltiples dominios.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 4

- b. Exponga con mayor grado de detalle una taxonomía acerca del ASBA donde ubiques la propuesta de solución y donde se delimiten las áreas teóricas de investigación. Incluir en el análisis medidas relevantes de calidad de los métodos sobre los conjuntos evaluados en la investigación.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (4 a.)

## Respuesta Pregunta 4 a.

- a. En este sentido podría entenderse que el aprendizaje continuo es la alternativa que el AP no es capaz de cubrir para abordar problemas de esta naturaleza. Exponga elementos acerca de este enfoque, teniendo en cuenta otras variantes del AP como pueden ser los modelos de atención con estructuras recurrentes como seq2seq u otras arquitecturas evaluadas en escenario de múltiples dominios.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (4 a.)

## Respuesta Pregunta 4 a.

- El aprendizaje continuo (AC) es una alternativa para el aprendizaje en múltiples dominios, tareas o clases; según el problema a resolver.
- El AC **no es una alternativa** que excluye o resuelve un problema que no es capaz de resolver el AP.
- El AC utiliza modelos de AP u otros paradigmas para lograr nuevas propuestas en la solución de problemas (p.ej., componentes base, hibridación de componentes).
- En la investigación presentada se emplea:
  - Para el modelo de extracción de aspectos un CNN que es un modelo de AP.
  - Para el modelo de clasificación de aspectos se emplea BERT que es un modelo de AP.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (4 a.)

## Respuesta Pregunta 4 a.

- Los modelos de AC que usan el paradigma de aprendizaje con Aislamiento de parámetros, agregan nuevas arquitecturas a los modelos de AP o que usan redes neuronales (p.ej., **PAN**, **DAN**, **HAT**; ver Taxonomía de métodos de AC).
- Uno de los retos principales del AC es estimar el **olvido catastrófico** al aprender un conocimiento nuevo y preservar la información anteriormente aprendida.

Los modelos de atención con estructuras recurrentes como seq2seq u otras arquitecturas evaluadas en escenario de múltiples dominios estudiados, hasta nuestro conocimiento, no proponen una solución o medida de evaluación a este importante reto al aprender un nuevo conocimiento apartir de uno anterior.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (4 b.)

## Respuesta Pregunta 4 b.

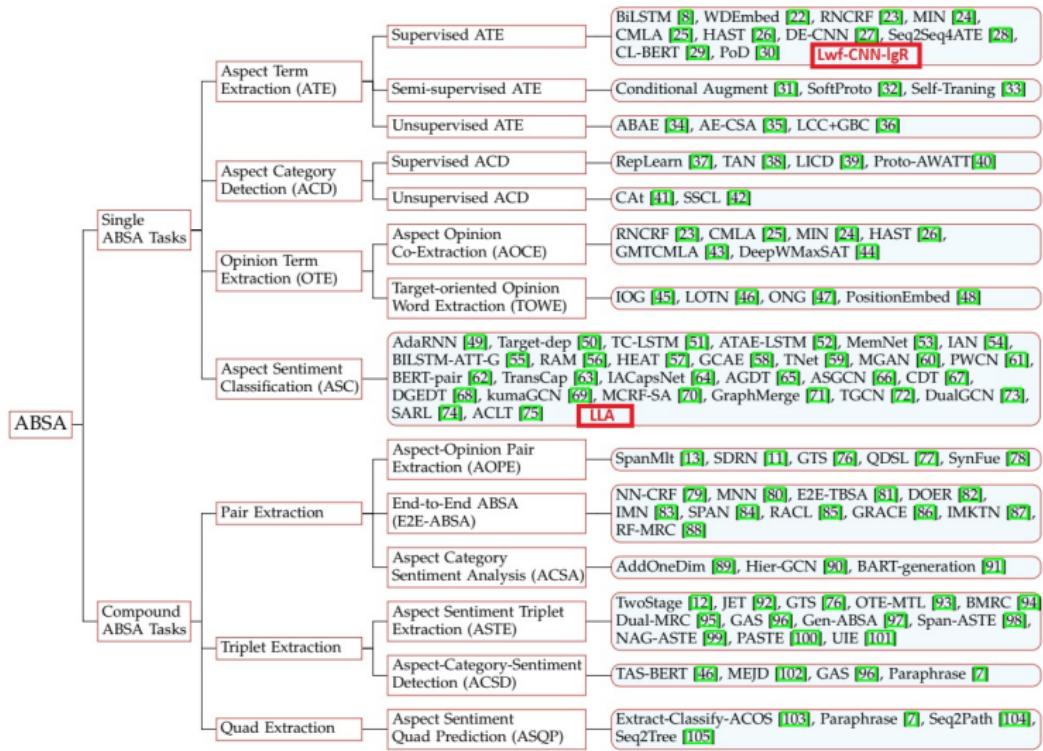
- b. Exponga con mayor grado de detalle una taxonomía acerca del ASBA donde ubiques la propuesta de solución y donde se delimiten las áreas teóricas de investigación. Incluir en el análisis medidas relevantes de calidad de los métodos sobre los conjuntos evaluados en la investigación.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (4 b.)

## Respuesta Pregunta 4 b.

Propuesta de taxonomía a partir del análisis del estado del arte propuesto en [Zhang, Dic. 2022].

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (4 b.)



# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 5

Al finalizar el capítulo 1 se expone como una de las conclusiones la siguiente: "Las propuestas de modelos que usan CNN no son muy empleadas para problemas de PLN, por los investigadores, debido a que esta estrategia se orienta a resolver problemas de procesamiento y clasificación de imágenes (por las características de la arquitectura de los modelos que la emplean). La adaptación de CNN a problemas de PLN es posible, pero incurren en el aumento del costo computacional (e.d., en memoria) al necesitar establecer un tamaño máximo para el texto presente en los ejemplos de los conjuntos de datos." :

- a) En tal sentido la primera propuesta de modelo computacional para la extracción de aspectos combina una CNN con AC. Explique en que se basó para formular esta variante.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) ( 5 a.)

## Respuesta Pregunta 5 a.

- Los valores de desempeño logrados por los investigadores en la propuesta de [Poria, 2016].

**Cuadro:** Conjunto de datos Semeval-2014 y de Efectos electrodomésticos (P:Exactitud, R:Exhaustividad y F1: F1-macro).

Conjuntos de datos	P	R	F1
Restaurante	88.27	86.10	87.17
Laptop	78.35	86.72	82.32
Efectos electrodomésticos	77.65	82.30	79.90

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) ( 5 a.)

## Respuesta Pregunta 5 a.

- Los estados del arte consultados referencian como una de las soluciones de mejores resultados la propuesta en [Poria, 2016] (p.ej., En el estado del arte propuesto en [Linan, Jul. 2022])

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) ( 5 a.)

## Respuesta Pregunta 5 a. (Continuación)

- Los investigadores que desarrollan la propuesta son de los más prolíficos en el área del análisis de sentimiento donde destaca **Erick Cambria** y su grupo de investigación de la Universidad de Singapore.
- Esta herramienta combina el uso de Aprendizaje Profundo (CNN) con reglas lingüísticas.
- La referencia como parte de los recursos externos en apoyo de las reglas lingüísticas de **SenticNet**; siendo un recurso multilingüe (incluye el idioma español)

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 5 (Continuación I)

- b. b. Un modelo de atención con una arquitectura bien formalizada que tenga en cuenta los diferentes dominios (por ejemplo, el aprendizaje multitarea o los modelos de atención) no suprimiría la necesidad de usar AC en la propuesta. Argumente.
- b. Podrías explicar de forma más detallada cada una de las etapas del modelo de extracción de aspectos en múltiples dominios con énfasis en la formalización matemática y demostración de la solución al problema de optimización.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Respuesta Pregunta 5 b.

- b. Un modelo de atención con una arquitectura bien formalizada que tenga en cuenta los diferentes dominios (por ejemplo, el aprendizaje multitarea o los modelos de atención) no suprimiría la necesidad de usar AC en la propuesta. Argumente.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) ( 5 b.)

## Respuesta 5 b.

Existe similitud entre los modelos que siguen el paradigma multi-tarea y el AC porque ambos tratan de usar el conocimiento compartido entre tareas. En el multi-tarea:

- En lugar de optimizar una única tarea, este optimiza varias tareas simultáneamente. Si se considera varias tareas como una única gran tarea, esto reduce al objetivo de una optimización tradicional.
- No es el objetivo acumular conocimiento en el tiempo y no se emplea el concepto del aprendizaje continuo
- El uso de una optimización global no es eficiente en términos de tiempo y recursos (memoria).
- El AC intenta retener conocimiento para permitir el aprendizaje incremental con la ayuda del conocimiento de una tarea anterior.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (5 c.)

## Pregunta 5 c.

- c. Podrías explicar de forma más detallada cada una de las etapas del modelo de extracción de aspectos en múltiples dominios con énfasis en la formalización matemática y demostración de la solución al problema de optimización.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) ( 5 c.)

## Respuesta 5 c.

Ver anexo C “Diagrama del proceso de entrenamiento del modelo de extracción de aspectos”.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) ( 5 c.)

## Respuesta 5 c. (Formalización matemática)

El objetivo del modelo es encontrar el conjunto óptimo de parámetros  $\theta$  que reduce la función objetivo  $J(\theta)$  de la arquitectura de red propuesta como modelo para la extracción de aspectos en la tarea ABSA.

$$J(\theta) = \mathbb{E}_{f(x, \theta) \sim \hat{p}_{data}} L(f(x; \theta), y) = \frac{1}{n} \sum_{i=1}^n L(f(x_i; \theta), y_i) \quad (7)$$

Donde  $n$  es la cantidad de ejemplos del conjunto de entrenamiento,  $x$  representa el vector o ejemplos a entrenar y  $y$  su etiqueta o valor a inferir

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) ( 5 c.)

## Respuesta 5 c. (Formalización matemática)

La solución al problema de optimización se encuentra en tratar de minimizar el error esperado en la función objetivo  $J(\theta)$  conjunto de entrenamiento del modelo.

Al limitar la capacidad del modelo propuesto de la red neuronal agregando una función de penalización a los parámetros de la red neuronal  $\Omega(\theta)$  a la función objetivo  $J$ :

$$\tilde{J}(\theta) = J(\theta) + \alpha\Omega(\theta) \quad (8)$$

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) ( 5 c.)

## Respuesta 5 c. (Formalización matemática)

La función objetivo del modelo de extracción de aspecto propuesto, combina una función de pérdida con una función de regularización:

$$\tilde{J}(\theta) = J(\theta) + \alpha \sum_{k=1}^{t-1} J(\theta)_{old\ k} \quad (9)$$

Donde  $t - 1$  representa la cantidad de dominios aprendidos anteriormente y  $t > 2$  para ser un modelo de AC.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) 5 d.

## Pregunta 5

- d. En la función objetivo de la página 39 el término de regularización, tal y como ha sido plasmado, depende de los datos, lo cual es poco común debido a que el rol de la regularización es lograr generalización en el modelo a través de la simplificación del mismo y por lo tanto muy pocas veces depende de los datos. Podría dar más elementos sobre el regularizador empleado.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) ( 5 d.)

## Respuesta 5 d.

Función objetivo del modelo de extracción de aspectos:

$$\tilde{J}(\theta) = J(\theta) + \alpha \sum_{k=1}^{t-1} J(\theta)_{old\ k} \quad (10)$$

Donde  $t - 1$  representa la cantidad de dominios aprendidos anteriormente y  $t > 2$  para ser un modelo de AC.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 6

Exponga de manera más detallada y sintética la formalización matemática de los modelos (secciones 2.1.1 y 2.3.3) haciendo las correcciones necesarias a posibles imprecisiones, los ajustes a la notación y las demostraciones necesarias para resolver el problema de optimización planteado en cada caso. :

- a. Establezca las relaciones teóricas de esta formalización con la manera de presentar los algoritmos 1, 2 y 3 de modo que estos se puedan reproducir (Utilice una escritura en lenguaje de pseudocódigo).

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 6

- Para un conjunto de dominios (e.d., texto de opiniones)  $\{1, \dots, T\}$
- Donde cada dominio tiene un conjunto de entrenamiento  $D_t$  y este tiene el par  $(X_t, Y_t)$ .
- Los modelos (secciones 2.1.1 y 2.3.3) son entrenados de forma incrementalmente para optimizar una función de costo:

$$\tilde{J}(\theta) = J(\theta) + \alpha \Omega(\theta) \quad (11)$$

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 6

Para el modelo de extracción de aspectos *Lwf-CNN-IgR* (sección 2.1.1) se define su función objetivo y término de regularización como:

$$\tilde{J}(\theta) = J(\theta) + \alpha \sum_{k=1}^{t-1} J(\theta)_{old\ k} \quad (12)$$

$$\tilde{J}(\theta) = -\log(P(y|x; \theta)) + \alpha \sum_{k=1}^{t-1} -\log(P(y|x; \theta))_{old\ k} \quad (13)$$

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 6

Para el modelo de clasificación de aspectos *LLA* (sección 2.3.3) se define el término de regularización como:

$$\Omega(\theta) = \sum_{k=1}^m \Phi_k (\theta'_k - \theta_k)^2 \quad (14)$$

$$\Phi_k = \frac{\sum^h \Delta \mathcal{L}_t}{T_t^2 + \xi} \quad (15)$$

- $h$  : La cantidad de pasos de actualización en el lote.
- $T_t$  : Es el movimiento total del  $\theta_k$ .
- $\Delta \mathcal{L}_t$ : Ecuación de actualización de los pesos.
- $\xi$ : Constante para evitar la división por cero.

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez)

## Pregunta 6

Para el modelo de clasificación de aspectos *LLA* (sección 2.3.3) se define su función objetivo y término de regularización como:

$$\tilde{J}(\theta) = J(\theta) + \alpha \sum_{k=1}^m \Phi_k (\theta'_k - \theta_k)^2 \quad (16)$$

$$\tilde{J}(\theta) = -\log(P(y|x; \theta)) + \alpha \left( \frac{\sum_t^h \Delta \mathcal{L}_t}{T_t^2 + \xi} \right) (\theta'_k - \theta_k)^2 \quad (17)$$

# Respuesta a las preguntas de la oposición (Dr.C Héctor Raúl González Diez) (6 a.)

## Respuesta Pregunta 6 a.

- a. Establezca las relaciones teóricas de esta formalización con la manera de presentar los algoritmos 1, 2 y 3 de modo que estos se puedan reproducir (Utilice una escritura en lenguaje de pseudocódigo).

# Seudocódigo algoritmo 1

---

## Algorithm 2 Modelo Lwf-CNN-IgR

---

### Entrada:

$\Theta_s$  ▷ Parámetros compartidos del modelo CNN  
 $\Theta_p$  ▷ Parámetros de los dominios anteriores (capa de clasificación)  
 $X_c, Y_c$  ▷ Oraciones y aspectos etiquetados en el dominio  $c$  actual

### Salida:

$\Theta_s$  ▷ Modelo CNN entrenado para el domino  $c$   
 $\Theta_c$  ▷ La capa de salida del último dominio

- 1:  $\Theta_c = \text{Inicializar los pesos de la red aleatoriamente para } c = 1$
- 2: **for**  $x_i, y_i \leftarrow X_c, Y_c$  **do**
- 3:      $Y_p[c - 1] = \text{Obtener valores de modelo CNN para } c - 1$
- 4:      $y_c = \text{Obtener valores de modelo CNN}$
- 5:      $sumLastDomain = \sum_{k=1}^{c-1} \text{para los } Y_p[c - 1]$
- 6:      $Loss'_i = Loss_i(y_c, y_i) + \lambda_o sumLastDomain$
- 7: **end for**

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Pregunta 1

Según se expone en el epígrafe 2.1, el modelo propuesto está inspirado en la solución reportada en [133], y además se plantea que la arquitectura en siete capas de CNN incluida en ese modelo es la misma que la concebida en [133].

- a. La solución referenciada en [133] tiene más de 6 años de ser publicada y se pueden identificar varias soluciones (más recientes) que utilizan aprendizaje profundo (incluso CNN), entonces, ¿qué criterios fueron tenidos en cuenta para seleccionar esta solución como base, en ventaja de otros trabajos relacionados reportados?
- b. Considerando que la arquitectura CNN es parte fundamental del modelo Lwf-CNN-IgR propuesto para la extracción de aspecto, describa y fundamente las diferencias de su modelo con respecto a la solución reportada en [133]

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Pregunta 1 (Continuación)

Según se expone en el epígrafe 2.1, el modelo propuesto está inspirado en la solución reportada en [133], y además se plantea que la arquitectura en siete capas de CNN incluida en ese modelo es la misma que la concebida en [133].

- c) Explique el procesamiento que se realiza a las opiniones, y justifique las decisiones adoptadas sobre la representación de estas

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Respuesta Pregunta 1 a

- a. La solución referenciada en [133] tiene más de 6 años de ser publicada y se pueden identificar varias soluciones (más recientes) que utilizan aprendizaje profundo (incluso CNN), entonces, ¿qué criterios fueron tenidos en cuenta para seleccionar esta solución como base, en ventaja de otros trabajos relacionados reportados? (Ver 5 a. (Dr.C Hector Hernández)).

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Respuesta Pregunta 1 a

- Los valores de desempeño logrados por los investigadores en la propuesta de [Poria, 2016].

**Cuadro:** Conjunto de datos Semeval-2014 y de Efectos electrodomésticos (P:Exactitud, R:Exhaustividad y F1: F1-macro).

Conjuntos de datos	P	R	F1
Restaurante	88.27	86.10	87.17
Laptop	78.35	86.72	82.32
Efectos electrodomésticos	77.65	82.30	79.90

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Respuesta Pregunta 1 a

- Los estados del arte consultados referencia como una de las soluciones de mejores resultados la propuesta en [Poria, 2016] (p.ej., En el estado del arte propuesto en [Linan, Jul. 2022])

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Respuesta Pregunta 1 a (Continuación)

- Los investigadores que desarrollan la propuesta son de los más prolíficos en el área del análisis de sentimiento donde destaca **Erick Cambria** y su grupo de investigación de la Universidad de Singapore.
- Esta herramienta combina el uso de Aprendizaje Profundo (CNN) con reglas lingüísticas.
- La referencia como parte de los recursos externos en apoyo de las reglas lingüísticas de **SenticNet**; siendo un recurso multilingüe (incluye el idioma español)

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Pregunta 1

Según se expone en el epígrafe 2.1, el modelo propuesto está inspirado en la solución reportada en [133], y además se plantea que la arquitectura en siete capas de CNN incluida en ese modelo es la misma que la concebida en [133].

- b. Considerando que la arquitectura CNN es parte fundamental del modelo Lwf-CNN-IgR propuesto para la extracción de aspecto, describa y fundamente las diferencias de su modelo con respecto a la solución reportada en [133].

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (1 b.)

## Pregunta 1 b

Cuadro: Diferencias entre Lwf-CNN-IgR y el propuesto en [133] .

Modelo	Entrada	AC	Conjuntos de datos
Lwf-CNN-IgR	Vector de BERT	Si	Diferentes
Modelo en [133]	Word2Vec	No	Laptop/Rest

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Pregunta 2

En la gran mayoría de las definiciones reportadas en la literatura sobre las tareas de análisis de sentimiento basado en aspectos, se plantea que los aspectos están asociados a entidades, productos o servicios, incluso en la formación de Bing Liu las entidades y los aspectos son elementos diferentes, pero con dependencia. También en el epígrafe 1.1 de la tesis se hace referencia a esto. La integración de los dominios (contenidos de opinión sobre diferentes entidades), problema planteado a resolver en la tesis, provoca que se pierda la referencia a la entidad objetivo y, como consecuencia, el adecuado sentido del aspecto (derivando en ambigüedades), con mayor impacto desde el punto de vista práctico.

- a. ¿Cómo es que se tiene contemplado esta cuestión dentro del modelo de extracción de aspectos propuesto? Explique y ejemplifique.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (2 a.)

## Respuesta Pregunta 2 a

No se tiene contemplada porque en su solución hay otros problemas y retos del PLN fuera del objeto de esta investigación:

- Reconocimiento de entidades subordinadas a otras entidades.
- Análisis de correferencias.
- Análisis de dependencias sintácticas.

Esta investigación se centra en la extracción y clasificación de aspectos explícitos en oraciones (e.d., una palabra simple o una frase) y que se diferencian de las entidades nombradas. (**pag. 9**).

Por ejemplo, (...) se reconocen las entidades nombradas [46]; (...) mediante el modelo propuesto, se extraen las que son aspectos y no son entidades nombradas. (**pag 54,55**).

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (2 b.)

## Pregunta 2 b.

- b. En caso de no tenerlo contemplado ¿Cómo considera que se podría tratar esta cuestión dentro de la concepción planteada?.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (2 b.)

## Pregunta 2 b.

La propuesta de [62] presenta un modelo no supervisado de aprendizaje en múltiples dominios (...). Este modelo tiene como principal estructura computacional un grafo que almacena las relaciones entre aspectos, entidades nombradas y su contexto. Para encontrar estas relaciones y nuevos patrones se usa un conjunto de reglas lingüísticas. (**pag 12**)

Proponemos a otros investigadores de la comunidad científica:

- El estudio de los conceptos y modelos asociados a los grafos convolucionales.
- Explotar las funcionalidades del marco de trabajo y biblioteca spaCy:
  - ① Reconocimiento de entidades.
  - ② Árboles de dependencias sintácticas.
  - ③ Módulos para el inglés y español.
  - ④ Marco para el entrenamiento de nuevos modelos.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Pregunta 3

En los epígrafes 2.2 y 2.6 se describen los métodos adoptados para la evaluación de los modelos de extracción y clasificación de aspectos, respectivamente. Sin embargo, se identifican varias cuestiones que requieren explicaciones adicionales.

- a. ¿Por qué se consideró realizar la evaluación de ambos modelos utilizando colecciones de prueba diferentes, si en la práctica ambas tareas deben formar parte del flujo de trabajo secuencial que actúa sobre una misma fuente de información?.
- b. Explique los objetivos que se perseguían a la hora de definir los dos tipos de pruebas que se declaran para evaluar el modelo Lwf-CNN-IgR y en qué consiste las reglas lingüísticas que se emplearon y por qué de su inclusión.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Pregunta 3 (Continuación)

- c) En ambos experimentos se aplican los modelos sobre varias colecciones de prueba en un mismo proceso de extracción o clasificación de aspectos y cada una de esas colecciones posee su propio “estándar de oro”. ¿Cómo es que se computan las métricas de precisión, recall y F1 empleadas bajo estas circunstancias? ¿Por qué fueron omitidas las métricas de precisión y recall de la evaluación de la clasificación de aspectos? (**Fácil**).
- c) Las colecciones de prueba seleccionadas, en su mayoría, son de las más usadas para evaluar este tipo de soluciones como se refiere en la tesis, por lo que es difícil entender por qué fueron considerados tan pocos trabajos en la comparación con la solución propuesta. Por tanto, extienda el análisis comparativo de la solución propuesta a otros trabajos reportados en la literatura y analice los resultados (**Difícil**).

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Pregunta 3

- a. ¿Por qué se consideró realizar la evaluación de ambos modelos utilizando colecciones de prueba diferentes, si en la práctica ambas tareas deben formar parte del flujo de trabajo secuencial que actúa sobre una misma fuente de información?.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (3 a.)

## Respuesta Pregunta 3 a.

Los conjuntos de datos empleados son:

- Para el entrenamiento del **modelo de extracción de aspectos** se uso:  
“Para evaluar el modelo propuesto, se seleccionaron **siete conjuntos de datos** de [62]; con opiniones en dominios como: reproductores de mp3, dvd, cámaras digitales y teléfonos inteligentes. Además, dos conjuntos de datos de [49] (e.d., uno sobre opiniones acerca de restaurantes y otro sobre laptops), y otro sobre opiniones de hoteles del sitio TripAdvisor”. (**pag 50**)

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (3 a.)

## Respuesta Pregunta 3 a.

- Para entrenar del modelo para la clasificación de aspectos se uso: "Para entrenar los modelos y evaluar su desempeño, los experimentos se realizaron empleando **siete de los conjuntos de datos** más usados en ABSA. Estos se describen en la Tabla 2.2 y fueron tomados de cuatro fuentes diferentes.  
El conjunto de opiniones sobre laptops y restaurantes fue tomado de (...) SemEval-2014 [49]; el conjunto de datos sobre dispositivos electrónicos fue tomado de [147, 148]; mientras que las opiniones sobre hoteles fueron tomadas de TripAdvisor [143]". (**pag 66**)

## Nota:

Se debió explicar mejor los conjuntos de datos en la especificación de los detalles de experimentación del modelo de extracción de aspectos

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (3 a.)

## Respuesta Pregunta 3 a. (Conjunto de Datos para reglas lingüísticas)

"Las reglas lingüísticas fueron aplicadas a un conjunto de datos no etiquetado, usado en [140] para aumentar el conjunto de entrenamiento, que este conjunto proporciona 1000 opiniones obtenidas desde Amazon sobre 50 tipos de dispositivos electrónicos, tales como teclados, tablets, etc". (**pag 50**)

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (3 b.)

## Respuesta pregunta 3 b.

- b. Explique los objetivos que se perseguían a la hora de definir los dos tipos de pruebas que se declaran para evaluar el modelo Lwf-CNN-IgR **y en qué consiste las reglas lingüísticas que se emplearon y por qué de su inclusión.**

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (3 b.)

## Respuesta pregunta 3 b.

Las dos pruebas fueron:

- Pruebas de dominio cruzado: Combinan seis conjuntos de datos etiquetados para el entrenamiento. Se realizan las pruebas en un séptimo dominio (no usado en el entrenamiento).
- Pruebas en el dominio: Se entrena y prueban los seis dominios, excluyendo un séptimo

Para comparar con los resultados de:

*"Lifelong-rl: Lifelong relaxation labeling for separating entities and aspects in opinion targets"*

**Autores:** L. Shu, B. Liu, H. Xu, and A. Kim.

**Publicada en:** "Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)", 2016, pp. 225–235

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (3 c.)

## Respuesta pregunta 3 c.

- c) En ambos experimentos se aplican los modelos sobre varias colecciones de prueba en un mismo proceso de extracción o clasificación de aspectos y cada una de esas colecciones posee su propio “estándar de oro”. ¿Cómo es que se computan las métricas de precisión, recall y F1 empleadas bajo estas circunstancias? ¿Por qué fueron omitidas las métricas de precisión y recall de la evaluación de la clasificación de aspectos?

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (3 c.)

## Respuesta pregunta 3 c.

Empleo de un contexto de aprendizaje incremental para diferentes dominios y la medida para estimar el olvido catastrófico,

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (3 d.)

## Respuesta pregunta 3 d.

- d. Las colecciones de prueba seleccionadas, en su mayoría, son de las más usadas para evaluar este tipo de soluciones como se refiere en la tesis, por lo que es difícil entender por qué fueron considerados tan pocos trabajos en la comparación con la solución propuesta. Por tanto, extienda el análisis comparativo de la solución propuesta a otros trabajos reportados en la literatura y analice los resultados.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (3 d.)

## Respuesta pregunta 3 d.

El modelo CLASSIC, con el que fue comparado el modelo LLA propuesto en mi investigación, fue comparado con **46 modelos del estado del arte**, según los autores [Ke, 2021], los cuales incluyen aprendizaje no continuo y continuo. En el caso de los métodos de **aprendizaje continuo estos son 38**.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (3 d.)

## Respuesta pregunta 3 d.

Hasta el momento de esta investigación se han reportado cuatro trabajos sobre AC y ABSA.

Esta revisión incluye CLASSIC y KAN, el cual es comparado en [Ke, 2021].

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (3 d.)

## Respuesta pregunta 3 d.

Para poder comparar la propuesta presentada en esta investigación con modelos de AP para la tarea ABSA es necesario:

- El modelo propuesto use algún método de aprendizaje continuo (p.ej., regularización, o aislamiento de parámetros)
- Los modelos sean entrenados en un entorno DIL (aprendizaje incremental de dominios).
- Presenten resultados para una medida de desempeño que evalue el olvido catastrófico.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Pregunta 4

En varios momentos de la descripción de la solución propuesta y las conclusiones del capítulo 2, se menciona el uso de “reglas lingüísticas”, con diferentes propósitos, por ejemplo: “... para la extracción de aspectos. . .”, “... para incrementar el conjunto de datos de entrenamiento”, entre otros usos, sin hacer mención o explicar:

- ¿Cuáles son estas reglas?
- ¿Cómo se definieron?
- ¿Cómo se usan?

entre otros elementos relevantes que permitan una adecuada comprensión de esta técnica y su aplicación. Por tanto, profundice en la explicación sobre concepción, propósito y uso de las reglas lingüísticas dentro de su propuesta.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Respuesta Pregunta 4 ¿Cuáles son estas reglas?

El nuevo modelo propuesto en esta investigación está inspirado en [Poria, 2016].

Las “reglas lingüísticas” fueron definidas por los autores de [Poria, 2016] y propuestas en el propio trabajo:

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Pregunta 4 Reglas lingüísticas

- **Regla 1:** Dada un sustantivo  $h$  sea un sujeto de una palabra  $t$ , el cual tiene un modificador adverbial o adjetivo presente en el diccionario de sentimiento, **SenticNet**. Entonces  $h$  es marcado como un aspecto.
- **Regla 2:** Excepto cuando las oraciones tengan un verbo auxiliar como, "is", "was", "would", "should", "could", "etc", se aplica:
  - **Regla 2.1:** Si el verbo  $t$  es modificado por un adjetivo o adverbio o está en una relación de modificación adverbial con otro token, entonces  $h$  es marcado como un aspecto (ver **Regla 1**). Eje: en "*The battery lasts little*", *battery* es el sujeto de *lasts*, el cual es modificado por un adjetivo modificador como *little*, entonces *battery* es marcado como un aspecto.
  - **Regla 2.2:** Si  $t$  tiene un objeto directo que es un sustantivo  $n$ , no encontrado en *SenticNet*, entonces  $n$  es marcado como aspecto, Eje: "I like the lens of this camera".

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Respuesta Pregunta 4 Reglas lingüísticas (Continuación )

- **Regla 3:** Si un sustantivo *h* es un complemento de un verbo copular, entonces *h* es marcado como un aspecto explícito, Eje: “The camera is nice”, **camara** es marcada como un aspecto.
- **Regla 4:** Si un término marcado como un **aspecto** por el CNN o las otras reglas está en una relación de composición sustantivo-sustantivo con otra palabra, en lugar de un término de aspecto de ambos. Eje Si en “battery life”, “battery” o “life” son marcados como un aspecto, entonces la expresión completa es marcada como un aspecto.
- **Regla 5:** Las reglas anteriores 1-4 aumentan la efectividad para descubrir más términos que son **aspectos** para aumentar la precisión se eliminan stop-words como *of*, *the*, *a*, etc, aunque sean marcados como términos y aspectos por el CNN u otras reglas.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Respuesta Pregunta 4 ¿Cómo se usan?

Inicialmente con el conjunto de reglas propuestas en [Poria, 2016] y la información presente en **Senticnet** se aprenden nuevos posibles tokens usando un conjunto de entrenamiento diferente al del modelo CNN.

Durante el proceso de evaluación del entrenamiento del modelo CNN se combinan las reglas con el CNN de la siguiente forma:

Las reglas y CNN se ejecutan en paralelo, entonces los términos marcados por cualquiera de los clasificadores son reportados como aspectos, excepto para los desmarcados por la última regla (Regla 5).

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Respuesta Pregunta 4 ¿Cómo se definieron?

Aunque presentadas en [Poria, 2016] son un subconjunto de las reglas propuesta en [Qiu, 2011] y mejoran los resultado al emplear un lexicón con mayor información como *Senticnet*.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Respuesta Pregunta 4 propósito y uso de las reglas lingüísticas

- Aumentar el conjunto de posibles aspectos a extraer.
- Son una herramienta computacional para aumentar el conjunto de entrenamiento.
- El uso de las reglas para aumentar el conjunto de entrenamiento y el posterior uso de modelos de lenguajes (p.ej., BERT, Gpt, T5) que tiene en cuenta en su entrenamiento el contexto donde ocurre el término o aspecto.
- En la investigación propuesta se modificó el uso del Standford Parser por la herramienta de análisis de dependencia para inglés de spaCy.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Pregunta 5

El problema planteado en la tesis, así como el objetivo general, se trata “...baja efectividad de la clasificación de aspectos en múltiples dominios en idioma inglés...”, sin embargo, los resultados experimentales indican que mientras más diversidad hay en los dominios menos eficacia se logra con la solución propuesta.

- a. ¿Cuáles son sus consideraciones respecto a este comportamiento?
- b. En la práctica los analistas o decisores no tienen por qué disponer de herramientas para llevar a cabo los análisis de similitud correspondientes, entonces, ¿Qué recomienda hacer para reducir esa dependencia de la solución propuesta a que las colecciones de entrada sean similares semánticamente?

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Pregunta 5

El problema planteado en la tesis, así como el objetivo general, se trata “...baja efectividad de la clasificación de aspectos en múltiples dominios en idioma inglés...”, sin embargo, los resultados experimentales indican que mientras más diversidad hay en los dominios menos eficacia se logra con la solución propuesta.

- c) ¿Cuál es la argumentación bajo la cual se hace la acotación de que su solución está concebida para el idioma inglés, y en consecuencia recomienda extender el estudio a propuestas de modelos para el idioma español?

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (5 a.)

## Respuesta Pregunta 5 a.

a) ¿Cuáles son sus consideraciones respecto a este comportamiento?

- Evitar o reducir el olvido catastrófico es difícil y campo de investigaciones futuras.
- Es necesario la creación de una medida que permita la estimación del olvido catastrófico de manera estándar para varias tareas imágenes, PLN, etc.
- Los modelos computacionales reflejan el mismo comportamiento que el aprendizaje humano al aprender tareas muy diferentes donde no existe referentes anteriores (p.ej.; medicina y aeronáutica).
- El uso de modelos computacionales para el aprendizaje continuo de dominios muy diferentes puede ser costoso y no obtenerse buenos resultados.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (5 b.)

## Respuesta Pregunta 5 b.

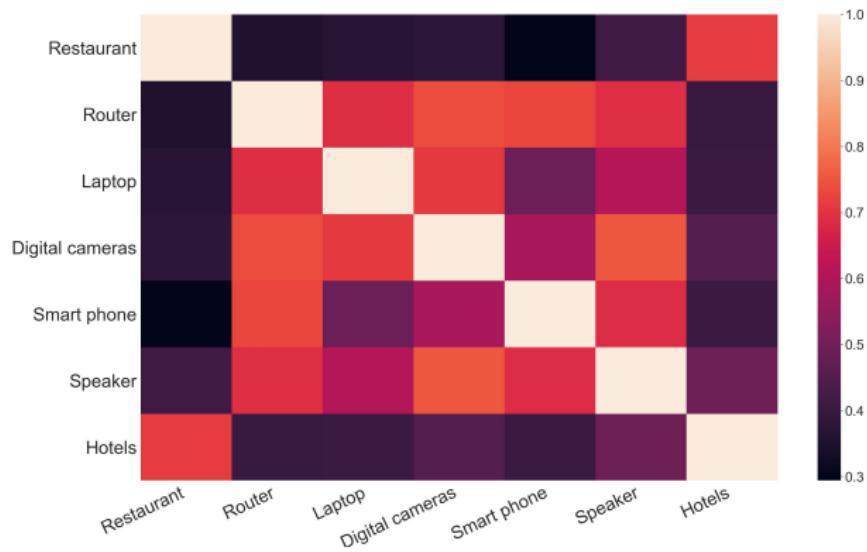
- b. En la práctica los analistas o decisores no tienen por qué disponer de herramientas para llevar a cabo los análisis de similitud correspondientes, entonces, ¿Qué recomienda hacer para reducir esa dependencia de la solución propuesta a que las colecciones de entrada sean similares semánticamente?

Consideramos recomendable realizar un análisis “a priori” sobre la similitud de las tareas a resolver:

- A través de la consulta con especialistas.
- El análisis de posibles representantes o grupos de representantes.
- El análisis de la similitud de los dominios a ser aprendidos (p.ej.; mapa de calor de la cercanía semántica realizado en la investigación)

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (5 b.)

Figura: Semejanza consenso entre el centroide de cada dominio.



# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (5 c.)

## Respuesta Pregunta 5 c.

- e. ¿Cuál es la argumentación bajo la cual se hace la acotación de que su solución está concebida para el idioma inglés, y en consecuencia recomienda extender el estudio a propuestas de modelos para el idioma español?

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (5 c.)

## Respuesta Pregunta 5 c.

- c. El enfoque es similar a la pregunta 1.d.i, realizada por el Dr.C Héctor Raúl González Diez (respuesta a partir de los argumentos de la anterior).

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (5 c.)

## Respuesta Pregunta 1 d.

Algunos elementos son:

- ① La exigencia de la investigación doctoral de demostrar el empleo práctico del modelo teórico propuesto.
- ② BETO[Cañete, 2020] es el primer pre-modelo basado en BERT para idioma español.
- ③ Han surgido varios pre-modelos, para el español, basados en BERT o con arquitectura de Transformers.
- ④ No se cuenta con muchos conjuntos de datos con las características de la tarea ABSA para el español:
  - Texto de la opinión.
  - Aspecto (palabra en el texto).
  - Polaridad del aspecto.
- ⑤ Varias de las instituciones consultadas mostraron su desinterés.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas)

## Pregunta 6

En el capítulo 3 de la tesis se tuvo la intención de presentar, a través de un caso de estudio, la aplicación práctica de la solución teórica propuesta, específicamente en el ámbito del procesamiento de opiniones generadas en una campaña de bien público. Sin embargo, hay mucha desproporción entre la información de contexto o motivación y lo relacionado con la aplicación concreta de la solución acompañado y sus resultados.

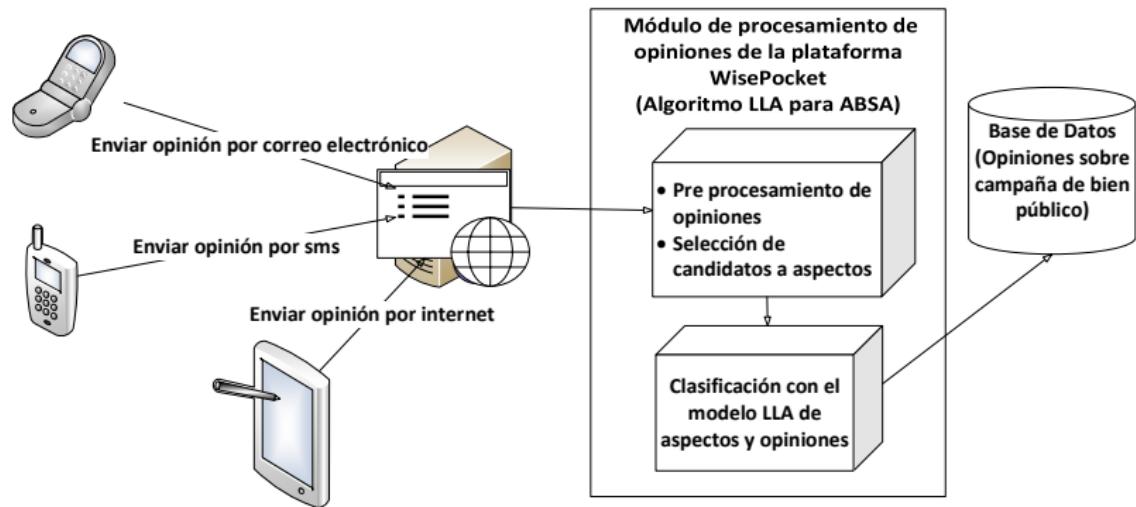
- a. Explique y ejemplifique como es que fue aplicada la solución teórica en ese escenario planteado.
- b. Realice un análisis detallado de los resultados obtenidos, resaltando dificultades que se presentaron en el proceso de aplicación de la propuesta, así como limitaciones detectadas de la propia propuesta.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (6 a.)

## Respuesta Pregunta 6 a.

- a. Explique y ejemplifique como es que fue aplicada la solución teórica en ese escenario planteado.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (6 a.)



**Figura:** Diagrama de despliegue del uso de los modelos *Learning without Forgetting with Linguistic Rules*; Lwf-CNN-IgR) y (*Lifelong Learning of Aspects*; LLA) en la plataforma WisePocket para la retroalimentación de las opiniones en la divulgación del **XIII Seminario Nacional sobre Estudios Canadienses**.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (6 a.)

## Respuesta Pregunta 6 a.

- a. Explicar Anexo F. Módulos de la plataforma Wisepocket (Informe de tesis)

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (6 b.)

## Pregunta 6 b.

- b. Realice un análisis detallado de los resultados obtenidos, resaltando dificultades que se presentaron en el proceso de aplicación de la propuesta, así como limitaciones detectadas de la propia propuesta.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (6 b.)

## Respuesta Pregunta 6 b.

**Cuadro:** Resultados obtenidos por el modelo de análisis de opiniones.

Opiniones	Aspectos	Positivos	Negativos	Neutros
50	120	50	30	40

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (6 b.)

## Respuesta Pregunta 6 b.

- En 7 opiniones el algoritmo dio una clasificación errónea (p.ej. "The conference was perfect but the coffee break was incredible" donde *perfect* fue clasificada como **positiva** e *incredible* **negativa**).
- Se realizaron 68 envíos de opiniones y en 11 de los casos el algoritmo de extracción no detectó los posibles aspectos (desechados como posibles opiniones).
- 7 opiniones fueron enviadas en español como si fueran opiniones en inglés (error humano).
- De los 11 casos el modelo de extracción no detectó 7 posibles opiniones y el resto de los casos fueron errores ortográficos enviados por los usuarios.

# Respuesta a las preguntas de la oposición (Dr.C Alfredo Javier Simón Cuevas) (6 b.)

## Consideraciones para el uso práctico de los modelos

- El uso de una herramienta de PLN para el reconocimiento de entidades y lograr relacionar los aspectos de forma efectiva a las entidades.
- Uso de una herramienta computacional para el reconocimiento del idioma inglés o español (según el caso).
- Uso de una herramienta para la corrección ortográfica del texto de las opiniones.
- Aumentar el entrenamiento de los modelos con conjuntos de datos que contengan un alto grado de lenguaje informal, teniendo en cuenta su posible empleo en el análisis de campañas de bien público.

# Modelos de Lenguaje: Fuentes de entrenamiento y parámetros

Cuadro: Fuentes de entrenamiento, conjuntos de datos y parámetros de pre-modelos para el PLN.

Modelo	Corpus	Parámetros
<i>BERT-base</i> (2019)	3.3B tokens (13GB data)	110M
<i>GPT-3</i> (2020)	500B tokens	175B
<i>T5</i> (2020)	200B tokens (750GB data)	11B

# Referencias



Michael McCloskey and Neal J. Cohen(1989)

Catastrophic interference in connectionist networks: The sequential learning problem

*Psychology of learning and motivation* 24, 109 – 165.



Robert M. French(1999)

Catastrophic interference in connectionist networks: The sequential learning problem

*Trends in cognitive sciences* 3(4), 128 – 135.



Stephen T. Grossberg(2012)

Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control

*Book* 3(4), 70.

# Referencias II

 Gido M. Van de Ven, and Andreas S. Tolias(2019)

Three scenarios for continual learning

*arXiv preprint arXiv:1904.07734,*

 Magdalena Biesialska, and Katarzyna Biesialska, and Marta R. Costa-jussa (2020)

Continual Lifelong Learning in Natural Language Processing: A Survey

*Proceedings of the 28th International Conference on Computational Linguistics,*  
6523 – 6541.

 Matthias Delange, and Rahaf Aljundi, and Marc Masana, and Sarah Parisot, and

Xu Jia, and Ales Leonardis, and Greg Slabaugh, and Tinne Tuytelaars, (2021)

A continual learning survey: Defying forgetting in classification tasks

*IEEE Transactions on Pattern Analysis and Machine Intelligence*

 Bin Liu, (2020)

Sentiment analysis: Mining opinions, sentiments, and emotions

*Sentiment analysis: Mining opinions, sentiments, and emotions*, isbn: 1-108-78728-2

 Vincenzo Lomonaco, (2019)

Continual Learning with Deep Architectures

*Thesis*, Universidad de Bologna

# Referencias III

 Gido M. Van de Ven, and Andreas S. Tolias(2019)

Three scenarios for continual learning

*arXiv preprint arXiv:1904.07734,*

 Jacob Devlin, and Ming-Wei Chang, and Kenton Lee, and Kristina Toutanova (2018)

Bert: Pre-training of deep bidirectional transformers for language understanding

*arXiv preprint arXiv:1810.04805,*

 Gido M. Van de Ven, and Andreas S. Tolias(2019)

Three scenarios for continual learning

*arXiv preprint arXiv:1904.07734,*

 Zhiyuan Chen, and Bing Liu (2018)

Lifelong machine learning

*Synthesis Lectures on Artificial Intelligence and Machine Learning,*

Book 12(3), 1–207

# Referencias IV

-  H. Shin, J. K. Lee, J. Kim, and J. Kim (2017)  
Continual learning with deep generative replay  
*Proceedings in NeurIPS*, 2990 – 2999.
-  S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert (2017)  
Continual learning with deep generative replay  
*Proceedings in NeurIPS*, 2001 – 2010.
-  Zhizhong Li, and Derek Hoiem (2018)  
Learning without forgetting  
*IEEE Transactions on Pattern Analysis and Machine Intelligence*,  
Book 40(12), 2935–2947
-  Zixuan Ke, and Bing Liu, and Hao Wang, and Lei Shu (2020)  
Continual Learning with Knowledge Transfer for Sentiment Classification  
*Proceedings of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*,  
Book 3, 683–698

# Referencias V



Zixuan Ke, and Bing Liu, and Hu Xu, and Lei Shu (2021)

Classic: Continual and contrastive learning of aspect sentiment classification tasks  
*arXiv preprint arXiv:2112.02714*,



Mehryar Mohri, and Afshin Rostamizadeh, and Ameet Talwalkar (2018)

Foundations of machine learning

MIT press,

Book



Yaqing Wang, and Quanming Yao, and James T. Kwok, and Lionel M. Ni (2020)

Generalizing from a few examples: A survey on few-shot learning

*ACM Computing Surveys CSUR*,

Book 53(3), 1–34



Soujanya Poria, and Erik Cambria, and Alexander Gelbukh (2016)

Aspect extraction for opinion mining with a deep convolutional neural network

*Knowledge-Based Systems*,

Journal 108, 42–49

# Referencias VI



Zhizhong Li, and Derek Hoiem (2018)

Learning without forgetting

*IEEE Transactions on Pattern Analysis and Machine Intelligence, Journal*, 2935–2947



Maria Pontiki, and Dimitris Galanis, and John Pavlopoulos, Harris Papageorgiou, and Ion Androutsopoulos, and Suresh Manandhar(2014)

Semeval-2014 task 4: aspect based sentiment analysis

*Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014),*

*Proceedings*, 27–35



Lei Shu, and Bing Liu, and Hu Xu, and Annice Kim(2016)

Lifelong-rl: Lifelong relaxation labeling for separating entities and aspects in opinion targets

*Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP),*

*Proceedings*, 225–235

# Referencias VII



Zhiyuan Chen, and Bing Liu (2014)

Topic modeling using topics from many domains, lifelong learning and big data  
*International Conference on Machine Learning,*  
*Proceedings*, 703–711



K.R Chowdhary (2020)

Natural language processing  
*Fundamentals of artificial intelligence,*  
*Journal*, 603–649



Ting Chen, and Simon Kornblith, and Mohammad Norouzi, and Geoffrey Hinton(2020)

A simple framework for contrastive learning of visual representations  
*Procedding in International conference on machine learning,*  
*Conference*, 1597–1607

# Referencias VIII



Arslan Chaudhry, and Puneet K. Dokania, and Thalaiyasingam Ajanthan, and Philip Torr(2018)

Riemannian walk for incremental learning: Understanding forgetting and intransigence

*Proceedings of the European Conference on Computer Vision (ECCV), Conference*, 532–547



Melanie Babst, and Thérèse Roux, and Johan Jager (2020)

Measuring the effectiveness of out-of-home advertising campaigns in South Africa  
*Communicare: Journal for Communication Sciences in Southern Africa, Journal*, 33–55



Baekkwan Park, and Kevin Greene, and Michael Colaresi (2020)

How to teach machines to read human rights reports and identify judgments at scale

*Journal of Human Rights, Journal 1(19)*, 99–116

# Referencias IX



Erik Cambria, and Poria Soujanya, and Devamanyu Hazarika, and Kenneth Kwok (2018)

SenticNet 5: discovering conceptual primitives for sentiment analysis by means of context embeddings

*Thirty-Second AAAI Conference on Artificial Intelligence, Conference, 1795–1802*



Bouras Dalila, and Amroune Mohamed, and Hakim Bendjanna (2018)

A Review of Recent Aspect Extraction Techniques for Opinion Mining Systems

*Natural Language and Speech Processing (ICNLSP), 2018 2nd International Conference on, Conference, 1–6*



Dionis López, and Leticia Arco (2019)

Aprendizaje profundo para la extracción de aspectos en opiniones textuales

*Revista Cubana de Ciencias Informáticas, Journal, 105–145*



Hai Do, and PWC Prasad, and Angelika Maag, Abeer Alsadoon (2019)

Deep Learning for aspect-based sentiment analysis: a comparative review

*Expert Systems with Applications,*

# Referencias X

 Komang Wahyu Trisna, and Huang Jin Jie (2022)

Deep Learning for aspect-based sentiment analysis: a comparative review  
*Applied Artificial Intelligence, Journal*, 1–37

 Ian Goodfellow, and Yoshua Bengio, and Aaron Courville(2016)

Deep learning  
*MIT press Cambridge, Book*

 Minlie Huang, and Yeqian Wang, and Xiaoyan Zhu, and Li Zhao(2016)

Attention-based LSTM for aspect-level sentiment classification  
*Association for Computational Linguistics, Conference*, 606–615

# Referencias XI



Mi Zhang, and Tieyun Qian (2020)

Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis

*Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Conference, 3540–35495*



Junyoung Chung, and Caglar Gulcehre, and Kyunghyun Cho, and Yoshua Bengio(2015)

Gated feedback recurrent neural networks

*International Conference on Machine Learning, Conference, 2067–2075*



Zixuan Ke, and Bing Liu (2022)

Continual Learning of Natural Language Processing Tasks: A Survey  
*arXiv preprint arXiv:2211.12701 (2022)*,



Cañete, José and Chaperon, Gabriel and Fuentes, Rodrigo and Ho (2020)

Spanish pre-trained bert model and evaluation data

*Practical Machine Learning for Developing Countries: learning under limited/low resource scenarios (PML4DC @ ICLR 2020), 1–10*

## Referencias XII



Akram, Arwa and Sabir, Aliea Salman (2022)

Explicit aspect extraction techniques

*Al-Qadisiyah for computer science and mathematics,*

*Journal*, 90–98



Zhang, Wenzuan and Li, Xin and Deng, Yang and Bing, Lidong and Lam, Wai (2022)

A survey on aspect-based sentiment analysis: tasks, methods, and challenges

*IEEE Transactions on Knowledge and Data Engineering,*

*Journal*, 1041-4347



Qiu, Guang and Liu, Bing and Bu, Jiajun and Chen, Chun (2011)

Opinion word expansion and target extraction through double propagation

*Computational linguistics,*

*Journal*, 37(1), 9–27

# Referencias XIV



Zhu, Linan and Xu, Minhao and Bao, Yinwei and Xu, Yifei and Kong, Xiangjie (2022)

Deep learning for aspect-based sentiment analysis: a review

*PeerJ Computer Science,*

*Journal, 2376-5992*