Revista Cubana de Ciencias Informáticas
ISSN: 1994-1536
ISSN: 2227-1899
Editorial Ediciones Futuro

López Ramos, Dionis; Arco García, Leticia
Aprendizaje profundo para la extracción de aspectos en opiniones textuales
Revista Cubana de Ciencias Informáticas, vol. 13, núm. 2, 2019, Abril-Junio, pp. 105-145
Editorial Ediciones Futuro

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

Tipo de artículo: Artículo original
Temática: Inteligencia Artificial
Recibido:08/08/2018 | Aceptado:03/04/2019

Aprendizaje profundo para la extracción de aspectos en opiniones textuales

Deep learning for aspect extraction in textual opinions

Dionis López Ramos[1,2*], Leticia Arco García[2,3]

[1] Departamento de Informática, Facultad de Ingeniería en Telecomunicaciones, Informática y Biomédica, Universidad de Oriente, Santiago de Cuba, Cuba. Avenida de Las Américas s/n, Santiago de Cuba, Cuba.
[2] Departamento de Ciencia de la Computación, Facultad de Matemática, Física y Computación, Universidad Central "Marta Abreu" de Las Villas. Carretera a Camajuaní km 5 ½ Santa Clara. Villa Clara. Cuba CP 54830.
[3] AI Lab, Computer Science Department, Vrije Universiteit Brussel, Pleinlaan 9, 1050 Brussels, Belgium.

[*]Autor para correspondencia: dionis@uo.edu.cu

Resumen

La extracción de aspectos en opiniones textuales es una tarea muy importante dentro del análisis de sentimientos o minería de opiniones, que permite lograr mayor exactitud al analizar la información y contribuir a la toma de decisiones. El aprendizaje profundo agrupa varios algoritmos o estrategias que han obtenido resultados relevantes en diversas tareas del procesamiento del lenguaje natural. Existen varios artículos de revisión sobre el análisis de sentimientos que abordan el aprendizaje profundo como una de las técnicas existentes para la extracción de aspectos; sin embargo, no existen artículos de revisión que se dediquen exclusivamente al empleo del aprendizaje profundo en el análisis de sentimiento. El objetivo de este artículo consiste en ofrecer un análisis crítico y comparativo de las principales propuestas y trabajos de revisión que emplean estrategias de aprendizaje profundo para la extracción de aspectos, profundizando en la forma de representación, modelos, resultados y conjuntos de datos empleados en esta tarea. En esta propuesta se hace el análisis de 89 artículos publicados durante el período 2011 a 2019 resaltando sus principales aciertos, fisuras, y retos de investigación. Finalmente, proponemos algunas direcciones de investigación futuras.

Palabras clave: Minería de Opiniones, Extracción de Aspectos, Aprendizaje Profundo, Procesamiento de Lenguaje Natural

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

Abstract

Aspect extraction in textual opinions is a very important task within the sentiment analysis or opinion mining, which allows achieving greater accuracy when analyzing information, and thus, contributing to decision making. Deep learning includes several algorithms or strategies that have obtained relevant results in various natural language processing tasks. There are several review papers on sentiment analysis that address deep learning as one of the existing techniques for extracting aspects; however, there are no review papers focus exclusively on to the use of deep learning in sentiment analysis. The main objective of this review paper is to offer a critical and comparative analysis of the main proposals and revision works that employ deep learning strategies for aspect extraction, by focusing on the representation approaches, principal models, obtained results and data sets used in experiments. In our proposal, the analysis of 53 papers published during the period 2011 to 2018 by highlighting their main successes, fissures, and research challenges is made. Finally, we propose some future research directions.

Keywords: *Opinion Mining, Aspect Extraction, Deep Learning, Natural Language Processing*

Introducción

Desde el comienzo del siglo XXI el análisis de sentimiento o minería de opiniones ha sido un área de investigación muy activa dentro del campo del procesamiento de lenguaje natural (*Natural Language Processing*; NLP) [1]–[3]. El estudio de las opiniones está asociado a áreas de la sociedad como la salud, el gobierno, la economía, entre otras. En los últimos años, han aumentado las propuestas comerciales para el uso del análisis de sentimientos. Compañías incipientes y grandes empresas como: Microsoft, Google, Hewlett-Packard y Adobe tienen sus propuestas para el análisis de opiniones sobre sus productos y servicios o los de terceros [4].

La minería de opiniones o el análisis de sentimientos es el estudio computacional de las opiniones, evaluaciones, actitudes y emociones que expresan las personas acerca de productos, servicios, organizaciones, individuos y eventos [4]–[6]. Las investigaciones y propuestas existentes tienen un alcance a nivel de documento, oraciones o aspectos (características). Aunque el análisis a nivel de documento y de oración es muy ventajoso en muchos casos, no garantiza obtener mucha información sobre el objeto de la opinión. Para realizar un análisis a mayor profundidad de las opiniones es necesario trabajar a nivel de aspectos.

El análisis de sentimientos a nivel de aspectos permite tener un mayor detalle de los sentimientos expresados por el autor o autores del texto analizado. Para lograr efectividad en este tipo de análisis se requiere primeramente de la identificación de las entidades presentes en la parte de la oración analizada (frase, oración o párrafo). A estas entidades se le deben asociar los aspectos que las describen y posteriormente clasificar los sentimientos asociados a este conjunto de información (entidades, aspectos). Ejemplos de posibles entidades son productos, servicios, personas, organizaciones

y eventos [7]. Las opiniones expresadas sobre los aspectos o características de una entidad pueden ofrecer mayor información. Por ejemplo, en una opinión sobre un producto, la persona que escribe una opinión expresa datos positivos o negativos sobre las características que lo conforman.

El análisis de sentimientos a nivel de aspectos es una tarea compleja porque además de reconocer los aspectos es necesario extraer la entidad a la cual están asociados. La extracción de entidades en el análisis de sentimiento es similar al problema clásico de reconocimiento de entidades nombradas (*Named Entity Recognition*; NER) [8], [9].

Las entidades se refieren a los nombres de productos, servicios, individuos, eventos y organizaciones y los aspectos se refieren a los atributos y componentes de estas entidades. Esta tarea presenta complejidades porque se debe identificar la categoría a la que pertenece la entidad (lugar, organización, evento, etc.) y también relacionar las posibles referencias a una misma entidad en un texto. Este última subtarea del análisis de entidades tiene el reto de resolver la posible sinonimia (varios nombres para identificar una misma entidad) y la polisemia (el nombre de una entidad tiene distintos significados), estando estrechamente relacionada a otras tareas del NLP como el análisis de correferencias [10], [11].

Los aspectos pueden estar representados por diferentes palabras o sinónimos. El uso de modificadores de opinión en uno o varios aspectos puede variar el sentimiento expresado en ellos [1]. Algunos autores [12]–[14] han tratado de enfrentar esta tarea a través de reglas lingüísticas o diccionarios de palabras, pero la construcción de estos recursos es muy costosa por el tiempo, personal y otros recursos que demanda. Una alternativa para lograr la extracción y clasificación de aspectos ha sido el empleo de diferentes modelos de máquinas de aprendizaje sobre grandes conjuntos de entrenamiento de dominios específicos (opiniones sobre: restaurantes, efectos electrodomésticos, etc.). La construcción de estos conjuntos puede ser cara y se incrementa cuando se desea extender a otros dominios o idiomas como el Español [4]. Una de las posibles soluciones es encontrar modelos que permitan aprender de varios conjuntos de datos y reconocer de forma automática reglas o patrones lingüísticos que puedan ser extensibles a otros dominios del conocimiento o idiomas [15]. Este trabajo está orientado a analizar las principales propuestas que usan aprendizaje profundo para la extracción de aspectos y la clasificación del sentimiento presente en ellos.

El análisis de sentimientos a nivel de aspecto es conocido en la literatura como Análisis de Sentimiento Basado en Aspectos (*Aspect Based Sentiment Analysis*; ABSA). Esta tarea fue llamada inicialmente análisis de sentimiento basado en características [1], [16] y está formada por dos subtareas principales:

- *Extracción de aspectos*: Se encarga de extraer aspectos y entidades de los documentos teniendo en cuenta que los aspectos pueden ser explícitos o implícitos. Por ejemplo, en la oración “La comida de ese restaurante es deliciosa y barata” se debe extraer como aspecto “la comida” de la entidad “restaurante”. En este ejemplo, el aspecto aparece como una palabra simple, pero este puede contener frases compuestas que hacen esta tarea aún

más compleja. Es importante tener en cuenta que la extracción de aspectos siempre está relacionada con una entidad [17]. En esta oración el adjetivo "barata" expresa la existencia de otro tipo de aspecto implícito que es el "precio". Algunos autores han trabajado el reto de identificar aspectos implícitos [18]–[20], aunque no abunda la literatura sobre este tema.

- *Clasificación de los sentimientos del aspecto*: Determina si la opinión que se emite sobre el aspecto es positiva, negativa o neutral. En la oración del ejemplo anterior la opinión acerca de la comida del restaurante es positiva.

En la propuesta de [21] se establecen tres importantes subtareas para el ABSA:

- Extracción del objeto de la opinión (*Opinion Target Expression*; OTE): Esta subtarea tiene como objetivo la extracción de los términos del aspecto (e.g., entidad o atributo).
- Detección de la categoría del Aspecto (*Aspect Category*; AC): Esta subtarea se relaciona con la identificación y agrupamiento de los aspectos en conceptos más generales como comida, confort, limpieza, etc.
- Polaridad del sentimiento (*Sentiment Polarity*; SP): Esta subtarea es encargada de asignar un sentimiento a los aspectos extraídos.

Por ejemplo, en la oración "El teléfono tiene una cámara potente pero una batería muy mala." los aspectos *cámara* y *batería* serían obtenidos por la subtarea OTE. En el caso de la subtarea AC estos aspectos pudieran ser clasificados como accesorios u otro concepto que pudiera agruparlos y la subtarea SP daría al aspecto *cámara* una polaridad positiva y a la *batería* negativa. La clasificación propuesta por [21] propone una mayor granularidad para el ABSA y varios de los trabajos analizados en esta revisión sistemática se relacionan con una o varias de estas subtareas.

Uno de los objetivos de la tarea ABSA es poder dar a un flujo de información no estructurado una estructura o forma de representación [22]. El Lenguaje de Marcación de Votación Universal (*Universal Voting Markup Language*; UVML) [23], es una propuesta para anotar con etiquetas las opiniones, pero no permite explotar la riqueza semántica que aparece relacionada a la opinión como son los aspectos. Los emoticons es una popular forma de asociar opiniones a una posible estructura a través de iconos o ideogramas [24]**,** pero la simpleza de esta forma de etiquetar no permite incluir los aspectos o características. En las propuestas [25], [26] se emplean el árbol sintáctico y las etiquetas morfológicas relacionadas a los aspectos y palabras de opinión que permiten obtener una estructura pero es difícil relacionar en ella la entidad a la que pertenecen los aspectos y otros datos de interés como el tiempo en que se emite la opinión.

Una forma de estructurar el texto de opiniones que explícitamente hace referencia a sus aspectos o características es la propuesta que se presenta en [4]. Ésta tiene en cuenta la relación entre entidades y aspectos, y considera la opinión como un quíntuplo: $O = (e, A, S, h, t)$, donde e es la entidad objetivo, A es el conjunto de aspectos de la entidad e, S es

el conjunto de sentimientos de la opinión asociado a los aspectos en *A*, *h* es quién emite la opinión y *t* indica cuándo se emitió la opinión. En el caso de los sentimientos que aparecen en *S*, éstos pueden ser positivos, negativos o neutrales, o establecer una escala de valores (1-5, cantidad de estrellas) u otra granularidad (positivos, muy positivos, neutros, negativos, muy negativos) que indique el grado de positividad o negatividad expresado [4], [6], [27].

Existen varias estrategias para la extracción de aspectos:

- Empleando la frecuencia con que aparecen los términos [28]–[31] .
- Analizando las relaciones sintácticas [19], [32]–[35].
- Usando técnicas de aprendizaje supervisado como Campos Aleatorios Condicionales (*Conditional Random Field*; CRF) [36], [37] y Máquinas de Vectores de Soporte (*Support Vector Machine*; SVM) [38] .
- Empleando métodos no supervisados basados en la detección de tópicos presentes en un documento según la Asignación Latente de Dirichlet (*Latent Dirichlet Allocation*; LDA) [39] y otras propuestas derivadas de esta [40]–[42].

Uno de los conceptos que ha tenido mucho éxito, al aplicarlo a varios dominios del conocimiento humano (procesamiento de imágenes, procesamiento de lenguaje natural, entre otros), es el aprendizaje profundo [43]. Las estrategias que siguen este concepto permiten el aprendizaje automático de las características de los datos de entrada en varias capas de abstracción y logran que el sistema aprenda las más complejas funciones. Las habilidades de aprender automáticamente cuáles características son importantes es fundamental cuando los datos tienen una alta dimensionalidad [43]. Algunas propuestas han sido aplicadas con éxito en la tarea de ABSA con buenos resultados. La existencia de muy pocos artículos de revisión que agrupen los principales resultados de ABSA empleando aprendizaje profundo motivó esta investigación. En este trabajo se hace un análisis de los principales resultados encontrados en la literatura que demuestran los casos exitosos empleando aprendizaje profundo para la extracción de aspectos. De ahí que el objetivo de este artículo de revisión consiste en ofrecer un análisis crítico y comparativo de las principales propuestas y trabajos de revisión que emplean estrategias de aprendizaje profundo para la extracción de aspectos, profundizando en las formas de representación, modelos, resultados y conjuntos de datos empleados en esta tarea.

Este artículo está organizado de la siguiente forma: en la sección “materiales y métodos” se explica cómo se realizó esta revisión sistemática de la literatura y se ofrecen respuestas a las diferentes preguntas de investigación relacionadas a los principales métodos, medidas de evaluación, principales conjuntos de datos de entrenamiento, etc. En la sección “resultados y discusión” se analizan los diferentes resultados de esta investigación. Finalmente, en las conclusiones se resumen las tendencias actuales del ABSA empleando métodos de Aprendizaje Profundo.

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

Materiales y métodos

Sobre el uso del aprendizaje profundo en el análisis de sentimientos se ha escrito una gran cantidad de trabajos, inicialmente enfocados en su mayoría a la detección de la polaridad de opiniones considerando todo el documento u oración donde son expresadas. Debido a los excelentes resultados reportados del uso de esta técnica, se han desarrollado y publicado varios artículos que proponen soluciones para la tarea ABSA empleando técnicas del aprendizaje profundo en el período de 2011 hasta 2019.

Para contribuir al avance en la tarea ABSA, en especial la extracción de aspectos empleando estrategias de aprendizaje profundo, es útil realizar una evaluación, identificación e interpretación de las investigaciones más relevantes hasta la fecha. Una búsqueda sobre revisiones de la literatura o estados del arte reveló la existencia de 27 estados del arte o revisiones sistemáticas de la literatura. Estas investigaciones tienen una escasa referencia a propuestas dirigidas a la tarea ABSA o es incompleta la referencia a soluciones que empleen técnicas de aprendizaje profundo.

En esta sección se muestran los criterios seguidos para realizar una revisión sistemática de la literatura (*Systematic Literature Review*; SLR) tomando como base las pautas propuestas en [44]. Se desarrolló una SLR sobre la extracción de aspectos empleando técnicas de aprendizaje profundo en el período de enero 2011 hasta febrero 2019. Si los estudios han sido publicados en más de una fuente o memorias de conferencias, se eligió el trabajo más completo. Se tuvieron en cuenta las publicaciones de varias fuentes de búsqueda de investigación científica: *ACM Digital Library*, *IEEE Explorer*, *ScienceDirect*, *Scopus*, *Springer Link* y *Google Scholar*. No se tomaron en cuenta en la investigación los trabajos que no tienen referencia de la revista, conferencia o memoria de evento, ni aquellos que son un resumen o publicación parcial de otro artículo. El área principal de investigación dentro de la cual pueden encontrarse artículos relevantes determina los términos principales de búsqueda. Los términos de búsqueda empleados fueron:

- “aspect extraction” + “opinion mining” + “deep learning”
- “aspect extraction” + “sentiment analysis” + “deep learning”
- “aspect-based sentiment analysis” + “deep learning”
- “aspect-level opinion mining” + “deep learning”

Las búsquedas realizadas permitieron obtener 53 artículos donde se aborda la extracción de aspectos utilizando aprendizaje profundo. Como muestra la Figura 1, a partir de 2015, aumentaron los trabajos sobre los métodos de aprendizaje profundo para la extracción de aspectos, esto se debe al surgimiento de herramientas y modelos que permiten entrenar eficientemente las propuestas basadas en aprendizaje profundo para resolver problemas del NLP, entre ellos el ABSA.

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

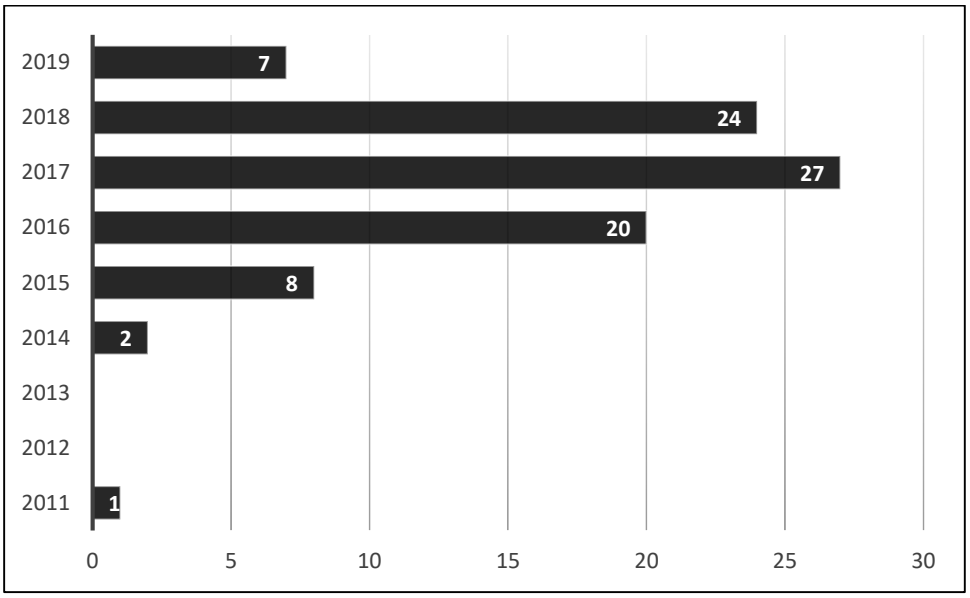

Figura 1. Cantidad de publicaciones sobre el uso del aprendizaje profundo en la extracción de aspectos en el período 2011- 2019.

En la Figura 1 se muestra como desde 2012 hasta 2013 no se encontraron publicaciones que relacionen el ABSA y las técnicas de aprendizaje profundo. Suponemos que la aparición de formas de representación del conocimiento más eficiente para el entrenamiento de redes neuronales en tareas del PLN como Word Embeddings propuesta por [45] y herramientas como word2vec [1] y Glove propuesta por [46] han permitido el aumento de trabajos que usen técnicas de aprendizaje profundo a partir de 2014.

Preguntas de investigación

En esta sección se presentan las preguntas de investigación que guían la SLR y se ofrece un resumen de los datos recogidos de los 89 artículos y 27 estados del arte para responderlas.

[1] https://code.google.com/archive/p/word2vec/

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

RQ1: ¿Los métodos de aprendizaje profundo para la tarea ABSA son tratados en los artículos de revisión?

Durante la investigación se encontraron más estados del arte y SLR sobre el análisis de sentimientos en general que específicamente sobre la tarea ABSA. En [27] se hace un análisis de varias estrategias que usan el aprendizaje profundo para el análisis de sentimientos mencionando con poca profundidad su empleo en la extracción de aspectos o características. Los autores sólo hacen alusión a tres artículos que abordan la tarea ABSA; aunque mencionan varias propuestas para la representación de las palabras y su uso con técnicas de aprendizaje profundo al enfrentar el análisis de sentimientos. El trabajo publicado en [47] tuvo como objetivo estudiar los métodos que realizan la extracción de aspectos pero sólo menciona uno que usa técnicas de aprendizaje profundo a pesar de mostrar algunas propuestas que las emplearon en otras tareas del análisis de sentimientos. En [48] se presenta un estudio del estado de arte de métodos que han dado una propuesta para la tarea ABSA pero no se menciona ninguno que empleara el aprendizaje profundo con tales propósitos. En [49] se abordan trabajos dirigidos al análisis de las palabras para determinar la emoción presente en las opiniones, pero no se relacionan propuestas para la extracción de aspectos y el uso del aprendizaje profundo. Los avances en el campo del análisis de opiniones para datos de diversos medios (fotos, texto, audio) y el empleo de algunas propuestas de aprendizaje profundo empleadas a tales efectos, aunque no especialmente dirigidas a la extracción de aspectos, son presentados en [50]. En el trabajo publicado en [51] se realiza un estudio de las propuesta para el análisis de sentimientos en citaciones científicas, mencionando la extracción de aspectos o características, y estableciendo como resultado que sólo existen dos propuestas que emplean el aprendizaje profundo. Una revisión sistemática de la literatura orientada a las propuestas existentes para el análisis de sentimientos usando dominios cruzados o adaptación al dominio es presentado en [52]. Esta investigación menciona la existencia de tres propuestas que emplean la extracción de aspectos empleando técnicas de aprendizaje profundo, pero esto resulta insuficiente para el conjunto de investigaciones sobre el tema. En [53] se ofrece un estudio de las propuestas existentes para el análisis de sentimientos en el dominio del Turismo. En artículo no se mencionan trabajos que emplean el aprendizaje profundo para la extracción o clasificación de aspectos. En [54] se presenta un interesante estado del arte sobre las propuestas asociadas a la determinación de las emociones del autor de textos de opinión. En esta investigación no se abordan soluciones o investigaciones asociadas a la extracción de aspectos, pero sí se hace referencia a algunas propuestas que emplean técnicas de aprendizaje profundo en el análisis de sentimientos. En [55] se analizan varias propuestas que realizan la extracción de aspectos, pero solamente se hace referencia a una investigación que realiza la tarea ABSA empleando técnicas de aprendizaje profundo. En [56] se presenta un estudio de las técnicas de aprendizaje profundo para el análisis de sentimientos, pero solo se menciona una propuesta específica sobre la extracción de aspectos. El estado del arte más

abarcador es el propuesto en [57]; no obstante, no se tuvieron en cuenta 23 artículos significativos sobre el tema. En [57] aparece un análisis de varias propuestas que usan el ABSA y estrategias de aprendizaje profundo. Se describen las principales características de estas propuestas; sin embargo, no se analiza cuál de ellas presenta el mejor resultado o las formas de representación del conocimiento y los conjuntos de datos más empleados, datos muy importantes para los investigadores.

Los 27 estados del arte o revisiones sistemáticas de la literatura que se consultaron en esta investigación hacen poca referencia a investigaciones que enfrentan la tarea de la extracción de aspectos usando algoritmos del aprendizaje profundo. Esto fundamenta la necesidad de realizar una investigación más profunda.

RQ2: ¿Cuáles son los métodos de aprendizaje profundo más empleados en la literatura?

Los modelos de aprendizaje profundo tienen varias mejoras sobre las máquinas de aprendizaje tradicionales. Dos de sus más destacados aportes son que reducen la necesidad de construir los datos (mediante un pre-procesamiento inicial) y realizar ingeniería de características en los conjuntos de entrenamiento [58]**.** Durante este proceso también pueden aparecer características no detectadas por los humanos.

Estos modelos consisten en técnicas de aprendizaje supervisado o no supervisado teniendo como estructura principal varias capas de Redes de Neuronas Artificiales (*Neural Networks*; NN) [58] que son capaces de aprender una representación jerárquica en arquitecturas profundas. Las arquitecturas de Aprendizaje Profundo están compuestas de varias capas de procesamiento, donde cada capa produce respuestas no-lineales basadas en la capa anterior y la entrada inicial.

Aunque las NN fueron introducidas en el siglo pasado, el crecimiento de las propuestas con Aprendizaje Profundo se enmarca en 2006 cuando Geoffrey Hinton presentó el concepto de Redes de Creencia Profunda [59]. El desarrollo de este tipo de modelo de aprendizaje ha sido posible por los avances del hardware en general y el desarrollo de Unidades de Procesamiento Gráfico (*Graphics Processing Unit*; GPU) y aceleradores de hardware. Además de la estructura de las redes neuronales, la profundidad de estos modelos y los avances del hardware, las técnicas de aprendizaje profundo han mejorado su desempeño a partir de:

- Uso de Unidades de Rectificación Lineal (*Rectified Linear Units*; ReLUs) como funciones de activación [60].
- Introducción de métodos *dropout* [61].
- Inicialización aleatoria de los pesos de las redes [62].
- Solución del problema de la desaparición del gradiente, así como su explosión con el uso de las redes de memoria de corto plazo (*Long Short-Term Memory*; LSTM) [63].

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

Arquitecturas de modelos de Aprendizaje Profundo

En esta sección se muestra una breve descripción de los modelos de aprendizaje profundo más comunes encontrados en este análisis sistemático de la literatura. Descripciones más detalladas sobre los modelos y arquitecturas de aprendizaje profundo se presentan en [58], [64]. La Tabla 1 resume estos modelos, atributos y características.

Tabla 1. Características principales de modelos de aprendizaje profundo.

Modelo	**Categoría**	**Tipo de aprendizaje**	**Entrada**	**Características**
Autoencoders (AE)	Generadores	No supervisado	Variado	• Adecuado para la extracción de características, reducción de dimensionalidad • El mismo número de unidades de entrada que de salida • Permite la reconstrucción de los datos de entrada en la capa de salida. • Trabaja con datos no etiquetados
Recurrent Neural Networks (RNN)	Discriminativa	Supervisado	Serial, series de tiempo	• Procesa una secuencia de datos a través de una memoria interna • Útil en problemas donde los datos dependen del tiempo (e.g., NLP, ABSA)
Long Short-Term Memory (LSTM)	Discriminativa	Supervisado	Serial, series de tiempo, datos de dimensión variable	• Buen desempeño con datos de tamaño variable y dependientes del tiempo • El acceso a las celdas de memoria es protegido por compuertas
Gated Recurrent Units (GRU)	Discriminativa	Supervisado	Serial, series de tiempo, datos de dimensión variable	• Buen desempeño con datos de tamaño variable y dependientes del tiempo • El acceso a las celdas de memoria es protegido por compuertas
Bidireccional RNN (e.g., BLSTM, BGRU, BRNN)	Discriminativa	Supervisado	Serial, series de tiempo, datos de dimensión variable	• Buen desempeño con datos de tamaño variable y dependientes del tiempo • El acceso a las celdas de memoria es protegido por compuertas
Convolutional Neural Networks (CNN)	Discriminativa	Supervisado	2-D (imágenes, sonidos, etc.), texto	• Las capas de convolución tienen la mayor parte del cálculo • Pocas conexiones entre capas, reduciendo el costo computacional • Necesitan de grandes conjuntos de datos para el entrenamiento
Restricted Boltzmann Machine (RBM)	Generativa	No supervisado, supervisado	Varios	• Adecuado para la extracción de características, reducción de dimensionalidad y clasificación • Costoso procedimiento de entrenamiento

En las redes neuronales el elemento más simple e importante es la neurona que puede recibir una o varias entradas y a través de su función de activación produce una salida. Cada neurona tiene un vector de pesos asociado al tamaño de la entrada y el sesgo que deben ser optimizados durante el proceso de entrenamiento. La salida es la entrada de la siguiente capa en la red neuronal. La capa final de la red representa la predicción del modelo. La función de pérdida determina la exactitud de esta predicción calculando el error entre los valores obtenidos en la última capa y los valores de comprobación. Un algoritmo de optimización como el Gradiente Descendente Estocástico (*Stochastic Gradient Descent*; SGD) [65] es empleado para ajustar los pesos de las neuronas calculando el gradiente de la función de pérdida. El índice de error es propagado hacia atrás en todos los pesos de las neuronas de la red (*Backpropagation* [66]). La red repite el ciclo de entrenamiento después de balancear los pesos de cada neurona en cada ciclo, hasta que el error alcanza una cota deseada. En el entrenamiento de las redes neuronales de los modelos de Aprendizaje Profundo hay dos importantes parámetros: el *epoch* y el *batch*. El *epoch* hace referencia a la cantidad de veces que es necesario pasar por los datos de entrenamiento para encontrar los pesos de la red que se ajustan mejor a los resultados esperados. El *batch* hace referencia a los subconjuntos de datos del entrenamiento que son tomados en cada iteración y evaluados. La función de pérdida se obtiene mediante el promedio de los valores de pérdida de los ejemplos pertenecientes al *batch*. Esta es una optimización que permite encontrar una solución rápidamente [58].

Los trabajos revisados emplean el aprendizaje profundo a través de enfoques supervisados, no supervisados o pueden usar propuestas híbridas combinando la salida de un método no supervisado con otro supervisado [43]. Algunos ejemplos de algoritmos de aprendizaje profundo expuestos en [43] y [58] y encontrados en el análisis realizado en esta investigación son: las Redes Neuronales Convolucionales (*Convolutional Neural Networks*; CNN) [67], las Redes Neuronales Recurrentes (*Recurrent Neural Network*; RNN) [66], la propuesta nombrada Memoria a Corto Plazo (*Long Short Term Memory*; LSTM) [63], las Unidades Recurrentes Cerradas (*Gated Recurrent Units*; GRU) [68] y sus variantes Bidireccionales, Autoencoders [69], [70], y las Máquinas de Boltzman Restringidas (*Restricted Boltzmann Machines*; RBM) [71]. A continuación, describiremos brevemente estas redes.

Redes Neuronales Convolucionales: Una CNN recibe una entrada en 2-Dimensiones (e.g., una imagen, una señal de audio, una oración). La capa de convolución es la parte principal de una CNN y consiste en un conjunto de parámetros a aprender, llamados filtros y que poseen la misma forma que la entrada, pero de menor dimensión. En el proceso de entrenamiento, el filtro de cada capa convolucional se mueve a través de todos los datos de entrada y calcula un producto interno entre la entrada y el filtro. Este cálculo permite mapear las características del filtro. Otra parte importante de la arquitectura de un CNN es la capa de asociación (*pooling*), la que opera sobre el mapa de características que se obtiene

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

como salida de la capa de convolución. El objetivo del *pooling* es reducir el tamaño espacial de la representación, con el objetivo de disminuir la cantidad de parámetros y el tiempo de cálculo y reducir la posibilidad de sobre entrenamiento. El *max pooling* es una estrategia que toma el valor máximo de cada región. Al resultado de la capa de convolución o *pooling* en una CNN se le aplica una función de activación. Es frecuente seleccionar ReLU, la cual consiste en neuronas con la función de activación de la forma $f(x) = max(0, x)$ aunque pueden se seleccionadas otras como la tangente hiperbólica [72]. La principal diferencia de CNN y las redes neuronales completamente conectadas es que cada neurona en CNN está conectada solamente con un pequeño subconjunto de la entrada. Esto disminuye la cantidad de parámetros en la red y disminuye el tiempo y la complejidad del entrenamiento [58]. La Figura 2 muestra una propuesta de arquitectura de una CNN.

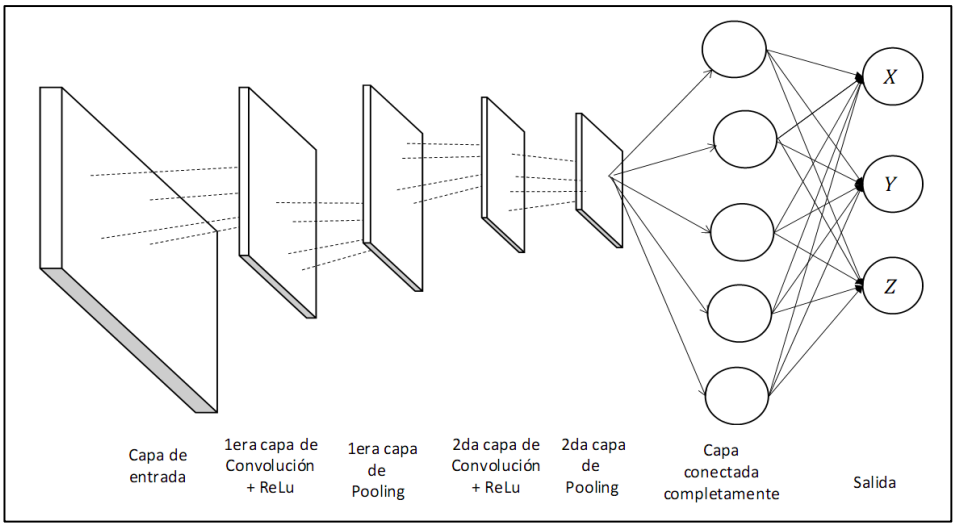

Figura 2. Arquitectura de una CNN de siete capas.

Redes Neuronales Recurrentes: Estas redes están orientadas a resolver problemas de secuencias o de series de tiempo (e.g., audio, texto) [64] que pueden tener tamaño variable. La entrada de una RNN consiste en el dato actual y el anterior. Esto significa que la salida en el instante $t - 1$ afecta la salida del instante t. Cada neurona está equipada con un ciclo de retroalimentación que retorna la salida actual como entrada del próximo paso, como se muestra en la Figura 3. Así, cada neurona en una RNN tiene una memoria interna que mantiene la información de los cálculos de la entrada anterior [43]. Una de las mayores debilidades de las RNN es el problema de la desaparición del gradiente (*vanishing gradient*), asociado a que éste sea cercano o igual a cero y el problema de la explosión del gradiente, que provoca que el gradiente aumente considerablemente [64].

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

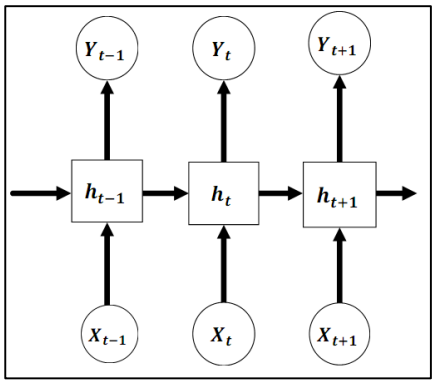

Figura 3. Arquitectura de una RNN de tres capas.

Memoria a Corto Plazo: LSTM es una extensión de las RNN y resuelve el problema de la desaparición y explosión del gradiente de las RNN. Estas usan el concepto de compuertas para sus neuronas. Cada una de estas compuertas calcula un valor entre 0 y 1 basado en su entrada. Incluyendo el mecanismo del ciclo de retroalimentación de una RNN para almacenar información, cada neurona en un LSTM (llamada una célula de memoria) tiene una compuerta multiplicativa de olvido (*forget gate*), lectura (*read gate*) y escritura (*write gate*). Estas compuertas son las que controlan el acceso de información en la neurona y evitan cualquier perturbación por datos de entrada irrelevantes. Cuando la compuerta de olvido está activa, la neurona escribe su dato dentro de ella. Cuando la compuerta de olvido está inactiva, la neurona olvida el contenido anterior. Cuando la compuerta de escritura es puesta a 1, otras neuronas pueden escribir en la neurona. Si la compuerta de lectura está activa, el contenido de la neurona pueda ser leído [73].

Unidades Recurrentes Cerradas: Las GRU **s**on una extensión de las RNN que evitan como las LSTM el problema de la desaparición y explosión del gradiente. La diferencia principal con respecto a las redes LSTM es que solamente definen dos compuertas: la de restauración (*reset gate*) y la de actualización (*update gate*) y manipula el flujo de información de manera similar a las redes LSTM sin la unidad de memoria. El uso de menos compuertas hace a la GRU menos costosa computacionalmente.

Redes Recurrentes Bidireccionales: Los tres modelos anteriores se enfocan en obtener el próximo estado a partir del anterior. Un propuesta que ha obtenido buenos resultados en modelos RNN es incorporar una capa hacia adelante y hacia atrás con el objetivo de aprender información de los tokens próximos y anteriores [74]. Como se muestra en la Figura 4, en cada instante t, una capa oculta hacia adelante $\vec{h}_t$ se calcula basándose en el estado previo $\vec{h}_{t-1}$ y la entrada actual x_t. De igual forma la capa oculta hacia atrás $\overleftarrow{h}_t$ se calcula basándose en el estado oculto futuro $\overleftarrow{h}_{t+1}$ y la entrada

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

actual x_t. La representación del contexto hacia adelante ($\vec{h}_t$) y hacia atrás ($\overleftarrow{h}_t$) son concatenados en un gran vector del instante t de la forma: $\overleftrightarrow{h} = [\overleftarrow{h}_t, \overleftarrow{h}_t]$.

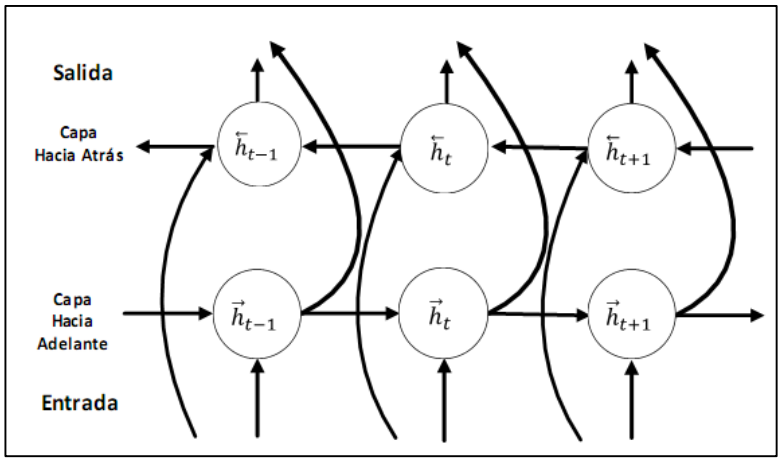

Figura 4. Arquitectura de una red recurrente bidireccional de tres capas.

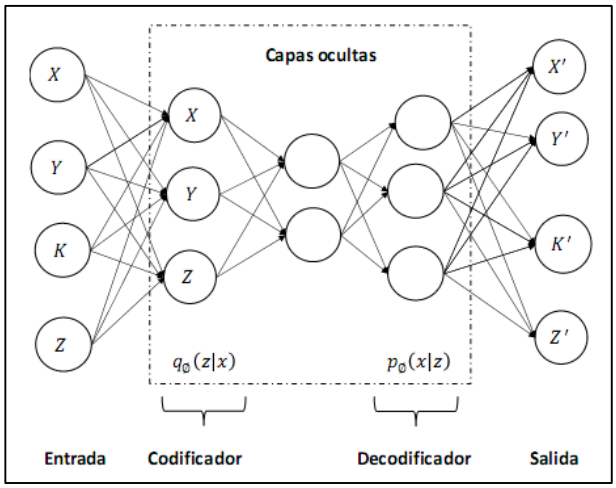

Figura 5. Estructura de un Autoencoder.

Autoencoders: Las redes AE consisten de una capa de entrada y una de salida que son conectadas a través de una o más capas ocultas, como se muestra en la Figura 5. Tienen la misma cantidad de neuronas de entrada que de salida. Estas redes tienen dos componentes principales: un codificador y un decodificador. El codificador recibe la entrada y trasforma ésta en una nueva representación, la cual es usualmente llamada un código o variable latente. El decodificador

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

recibe el código generado por el codificador y lo trasforma en una reconstrucción de la entrada original. El procedimiento de entrenamiento de estas redes tiene como objetivo la minimización del error de reconstrucción. Existen varias variaciones y extensiones de los AE, entre ellos los ruidosos, contractivos, apilados o variacionales [64].

Máquinas de Boltzman Restringidas: Una RBM es una red neuronal estocástica que consiste en dos capas: una capa visible que contiene la entrada conocida y una capa oculta que contiene las variables latentes. La reconstrucción de una RBM es aplicada a la conectividad de las neuronas comparadas con máquinas de Boltzman. Las RBM construyen un grafo bipartito, como se muestra en la Figura 6, de manera que las neuronas de la capa visible deben estar conectadas a todas las neuronas de la capa oculta, pero no existe conexión entre dos unidades de la misma capa. En estas redes el sesgo de la capa de entrada está conectado a todas las neuronas de esa capa (de igual forma ocurre con el sesgo en la capa oculta). Este tipo de red puede ser apilada para formar redes neuronales profundas y también puede crear bloques de redes de Creencia Profunda [58].

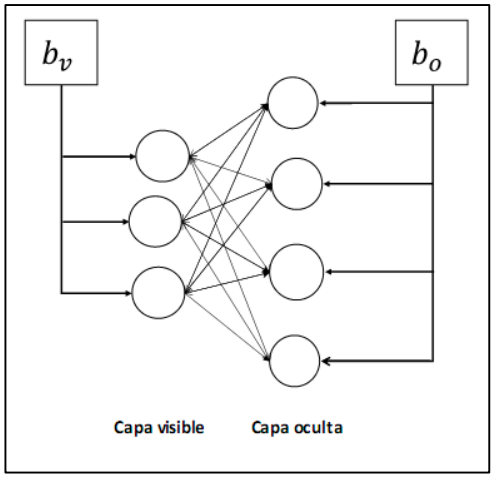

Figura 6. Estructura de una Máquina de Boltzman Restringida.

En el análisis realizado en esta investigación se encontraron más trabajos con modelos supervisados que no supervisados o híbridos. De las propuestas que siguen un enfoque supervisado, el 42% corresponden a variantes que usan LSTM, como se muestra en la Figura 7. La selección de este método está dado por la naturaleza secuencial de la información en las tareas de NLP para la información textual [58]. En los trabajos analizados se identificó que una de las características de las redes LSTM es su uso en forma bidireccional. Esta configuración de la red propone el entrenamiento de dos redes LSTM, donde la segunda recibe como entrada la salida de la primera, y ambas concatenan los estados ocultos [75]–[78].

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

En [26], [79], [80] se emplea una variante de CNN mediante el uso de una secuencia de redes convolucionales donde la salida de una red es la entrada de la otra. Este algoritmo es conocido como Redes Convolucionales Apiladas o en Cascada (*Convolucional Stacked Network*; CSN). La selección de este tipo de algoritmos por parte de los investigadores se justifica por la variedad de problemas de NLP que se pueden resolver aplicando las CNN y los buenos resultados que se han obtenido. Esta técnica permite, a partir de la representación de las palabras, aplicar en cada capa de la red una operación de convolución o de selección de características importantes.

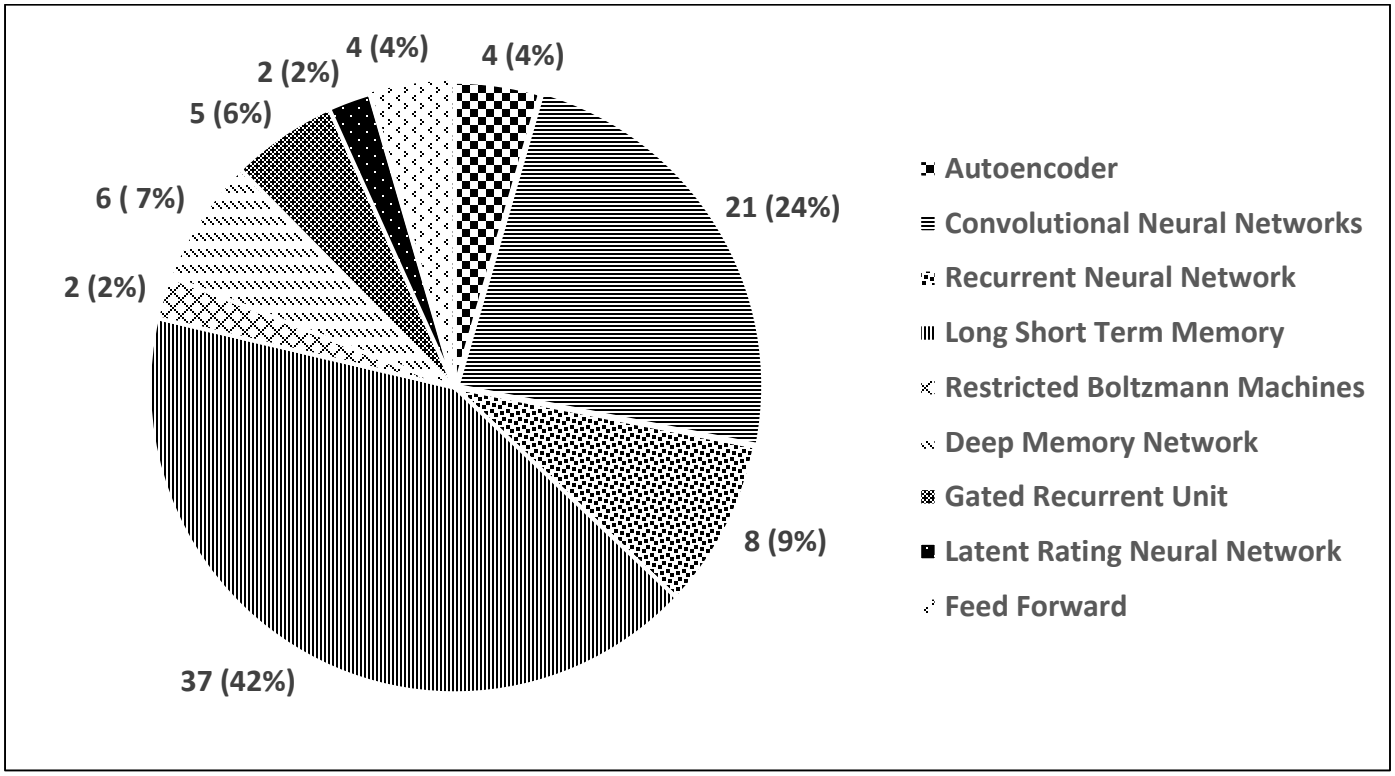

Figura 7. Clasificación de los métodos de aprendizaje profundo empleados en la extracción de aspectos, donde se muestra la cantidad de artículos y el porciento que representa del total.

Otros enfoques del aprendizaje profundo muy empleados de forma supervisada son GRU y RNN. El empleo de estos algoritmos por parte de los autores se justifica porque estas técnicas están especializadas para procesar secuencias de valores. Estos métodos procesan una oración desde el inicio hasta el final, analizando una palabra a la vez. Además, se auxilian de las relaciones de dependencias y los árboles sintácticos para extraer a nivel de palabras las relaciones semánticas y sintácticas. De esta forma, logran capturar las representaciones del conocimiento más abstractas y de más alto nivel en diferentes capas. Por otro lado, las RNN son capaces de modelar secuencias de tamaño arbitrario por la

aplicación de unidades recurrentes a lo largo de las secuencias de tokens. Las RNN tienen como desventajas el desvanecimiento o explosión del gradiente. Esto provoca que las RNN no sean suficientes para modelar dependencias de gran tamaño. Este problema ha motivado que varias propuestas usen las LSTM y GRU para la extracción de aspectos [81], [82].

El resto de los trabajos revisados emplean los algoritmos del aprendizaje profundo de forma no supervisada, como el uso de Autoencoders en la investigación publicada en [83]. Esta técnica implementa una red neuronal que copia los datos de la capa de entrada en la capa de salida. Internamente, tiene una capa oculta que se encarga de codificar y decodificar los datos de entrada a través de dos funciones. Por lo general, estas funciones se definen de forma que la copia sea aproximada y de esta manera el modelo es forzado a priorizar aquellos aspectos que sean propiedades útiles de los datos.

En [84] se prueba un RBM para la extracción de aspectos. Esta red neuronal es un modelo basado en energía con una distribución de probabilidad conjunta especificada por una función de energía. En [84] las unidades de la capa oculta representan aspectos, sentimientos previamente seleccionados y palabras de rechazo, mientras que la capa de entrada está asociada a las palabras de las oraciones de entrenamiento.

En [85], [86] se usa una Red de Memoria Profunda (*Deep Memory Network*; DMN) que es entrenada a partir de un conjunto de aspectos predefinidos. En varios trabajos [85]–[88] se reporta el empleo del Mecanismo de Atención (*Attention Mecanism*) que promedia los pesos que pueden ser relevantes en otros puntos de una red neuronal. Este mecanismo permite incluir características lingüísticas o sintácticas al proceso de aprendizaje de la red neuronal que implementa el algoritmo del aprendizaje profundo. Varias propuestas [74], [80], [89] agregan reglas lingüísticas al empleo de algoritmos del aprendizaje profundo.

Debido a que la extracción de aspectos es una tarea de gran importancia en el análisis de sentimientos, aparecen muchas propuestas que realizan la extracción de aspectos y la clasificación de la polaridad de los aspectos (positiva, negativa y neutra) de forma paralela [27], [90], [91].

En algunos trabajos analizados, la forma de realizar la hibridación es mezclando la salida de un método de aprendizaje profundo con el entrenamiento de una máquina de aprendizaje como CRF. En [92] primeramente se entrena una RNN, y la salida resultante se utiliza como entrada de un CRF. Esta propuesta no supera los resultados del método ganador en la competencia SemEval 2016 para la tarea ABSA. En [93] se relaciona la salida de un CNN con un CRF pero no mejora los resultados de SVM contra los que se prueba este método. Los resultados obtenidos en los artículos analizados indican que se debe revisar si realmente es beneficioso relacionar métodos de aprendizaje profundo con otros modelos, tales como CRF. En [94] se propone el uso de LSTM y en la salida de la red neuronal un CRF para el aprendizaje.

Empleando esta variante combinada de LSTM-CRF se obtuvieron resultados de más de un 83% de Micro-F1. En el caso de [53] es probado con los mejores métodos de la competición SemEval 2014 para la tarea ABSA y otros métodos que usan CRF. Para todos los casos los resultados de este método lo superan.

Algunas propuestas no tienen como objetivo extraer todos los aspectos asociados a una entidad, sino que tienen el propósito de extraer los aspectos que estén asociados a categorías prefijadas [72], [84], [90]. Por ejemplo, en [90] un conjunto de datos sobre restaurantes se evalúa para determinar cuán buena es la extracción de aspectos o palabras asociadas a comida, personal y ambiente. En la Figura 8 se muestra una taxonomía que clasifica los principales métodos de Aprendizaje Profundo analizados en esta investigación.

La clasificación correspondiente a los "Modelos de Aprendizaje Profundo" hace referencia a aquellos artículos que utilizan CNN, LSTM o GRU para la tarea ABSA [87], [89], [91]. La clasificación nombrada "Mecanismo de Atención + Modelos de Aprendizaje Profundo" engloba aquellos trabajos que combinan el Mecanismo de Atención con modelos como CNN, LSTM, BLSTM o BGRU [75], [77]. En esta investigación se encontraron 10 trabajos que realizan esta combinación y recientemente se ha convertido en una práctica muy extendida entre los investigadores en ABSA. La clasificación "Máquinas de Aprendizaje + Modelos de Aprendizaje Profundo" se refiere a los trabajos donde se usan modelos de Aprendizaje Profundo como CNN, LSTM o BLSTM y luego una máquina de aprendizaje de tipo CRF [92], [95].

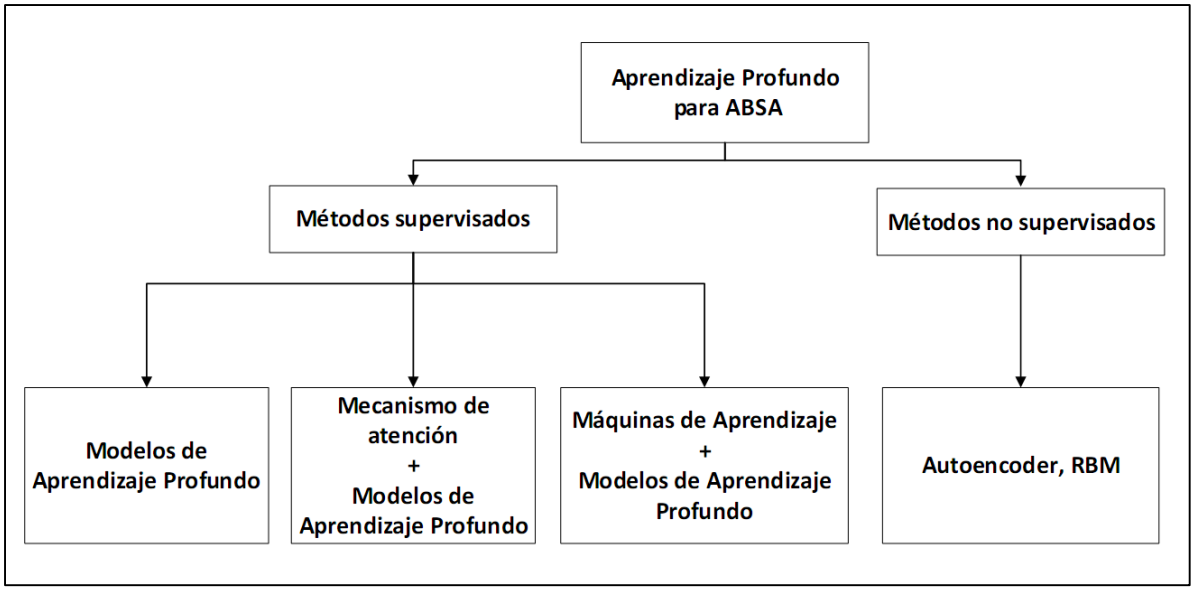

Figura 8. Taxonomía de métodos de Aprendizaje Profundo para ABSA.

Como se muestra en esta investigación los modelos más empleados para la extracción de aspectos son los supervisados. Estos modelos sufren la desventaja de necesitar para su entrenamiento muchos datos o ejemplos etiquetados [96]. La necesidad de usar grandes conjuntos de datos está asociada al uso de la regularización y el *dropout* para la reducción del error, así como el empleo del Gradiente Descendiente para encontrar los pesos de la red neuronal [58]. Por lo que, se debe seguir trabajando en la construcción e identificación de conjuntos de datos para ABSA teniendo en cuenta diversos dominios e idiomas [97]. Una posible solución para la existencia de pocos datos etiquetados para algunas clases o la aparición durante el entrenamiento de nuevas clases es el uso de estrategias como el aprendizaje *Few-shot* [94]. Este esquema de aprendizaje está en su fase de desarrollo pero las propuestas actuales permitirán a los investigadores en ABSA obtener mejores resultados con conjuntos de datos poco o no etiquetados [96], [98].

RQ3: ¿Cuáles son los conjuntos de datos empleados para el entrenamiento y evaluación de los métodos?

Los trabajos analizados han empleado diferentes conjuntos de datos para evaluar los métodos que realizan la extracción de aspectos, como se muestra en la Figura 9. Los conjuntos de datos más empleados son los de SemEval 2014, 2015 y 2016, Yelp, Amazon y SentiHood. "Ganu 2009" representa un conjunto de datos propuesto en [99] formado por opiniones sobre restaurantes y no se encuentra público. "MPQA"[2] es un conjunto de datos propuesto en [100] compuesto por noticias etiquetadas según las opiniones presente en ellas y "SSAC 2017" es el conjunto de datos del taller sobre detección de emociones, opiniones y análisis de sentimientos[3] celebrado en 2017. "Hu 2004" es el conjunto de datos propuesto en [1] compuesto por opiniones de efectos electrodomésticos y que se encuentra público para el uso de los investigadores.

En varios trabajos como en [25], [80], [101] se emplea el conjunto de datos propuesto en la competición SemEval 2014 [102] para evaluar la tarea ABSA. Este conjunto de datos está compuesto por oraciones que contienen opiniones sobre los dominios de restaurantes y laptops, como se muestra en la Tabla 2. Este conjunto posee para cada oración los aspectos presentes en ella y la opinión que se expresa sobre ellos. Para expresar la polaridad de la opinión asociada a los aspectos se establecieron las categorías: positiva, negativa, conflicto y neutral, como se muestra en la Tabla 3.

[2] http://mpqa.cs.pitt.edu/
[3] http://2017.eswc-conferences.org/

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

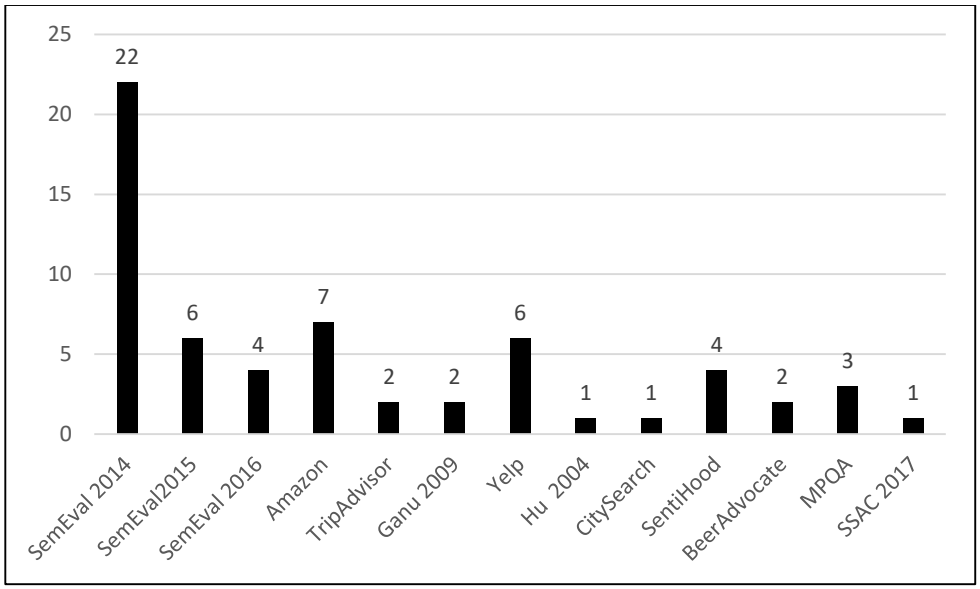

Figura 9. Uso en los trabajos analizados de las colecciones de opiniones disponibles para la validación**.**

Tabla 2. Cantidad de oraciones de opinión por dominios en el conjunto de datos de SemEval 2014.

Dominio	**Entrenamiento**	**Prueba**	**Total**
Restaurante	3041	800	3841
Laptop	3045	800	3845

Tabla 3. Cantidad de oraciones según la polaridad de los aspectos en el conjunto de datos de SemEval 2014.

Dominio	**Positiva**	**Negativa**	**Conflicto**	**Neutral**	**Total**
Restaurante	2892	1001	105	829	4827
Laptop	1328	994	61	629	3012

Los aspectos anotados en el conjunto de datos o corpus están agrupados por las categorías: comida, servicio, precio, ambiente (la atmósfera y ambiente en un restaurante), y anécdotas/misceláneas (no pertenecientes a ninguna de las anteriores). Para asignar la polaridad de cada término se tomó una ventana de seis palabras donde se buscaron los términos con polaridades más cercanas y se sumaron los valores asignados a cada término (-1: negativo, 1: positivo, 0: neutral). El tipo de polaridad “conflicto” se asigna cuando la suma es igual a 0 pero no todos los términos en la ventana son neutrales. La mayoría de los aspectos son palabras simples (2148 para el dominio de Laptops, 4827 para

restaurantes). Este conjunto está disponible en formato XML y se puede descargar a través de META-SHARE[4], repositorio dedicado a compartir recursos para el NLP.

Otro conjunto de datos usado en varias publicaciones [87], [103], es el propuesto en la competición SemEval 2015 [104], para evaluar propuestas en la tarea ABSA. Este conjunto de datos se caracteriza por tener oraciones en tres dominios (laptops, restaurantes y hoteles), aunque para el dominio hoteles no se proporcionó un conjunto de entrenamiento, como se muestra en la Tabla 4.

Tabla 4. Cantidad de oraciones de opinión por dominios en el conjunto de datos de SemEval 2015.

Dominio	**Entrenamiento**	**Prueba**	**Total**
Restaurante	1315	685	2000
Laptop	1739	761	2500
Hoteles		266	266

Al dominio laptops, de este conjunto de datos de SemEval 2015, se asignaron nueve categorías (general, precio, calidad, funcionamiento y rendimiento, usabilidad, diseño y características, portabilidad, conectividad, misceláneas). Fueron definidas cinco categorías para el dominio restaurantes (general, precio, calidad, estilo y opciones, misceláneas) y ocho para el dominio hoteles (general, precio, confort, precio, limpieza, calidad, diseño y características, misceláneas). La polaridad se define a nivel de oración teniendo en cuenta el documento donde aparece. El valor neutral de la polaridad se aplica cuando se tiene igual cantidad de términos positivos y negativos. En la Tabla 5 se muestra la cantidad de oraciones según la polaridad de los aspectos en el conjunto de datos de SemEval 2015. Este conjunto de datos, al igual que el conjunto de SemEval 2014, está disponible en formato XML a través de la plataforma META-SHARE.

Tabla 5. Cantidad de oraciones según la polaridad de los aspectos en el conjunto de datos de SemEval 2015.

Dominio	**Positiva**	**Negativa**	**Neutral**	**Total**
Restaurante	1320	601	79	2000
Laptop	1406	938	156	3012
Hoteles	191	66	9	266

[4] El conjunto de datos puede ser descargado de http://metashare.ilsp.gr:8080/. META-SHARE (http://www.meta-share.org/) fue implementado en el marco de trabajo de la red de excelencia META-NET (http://www.meta-net.eu/).

En [87], [105], [106] se usa el conjunto de datos propuesto en SemEval 2016 [104] para entrenar y evaluar los métodos propuestos. Este conjunto de datos consta de 39 subconjuntos, de éstos, 19 para entrenamiento y 20 para prueba. Los textos contienen información de siete dominios (laptop, teléfonos móviles, cámaras digitales, hoteles, restaurantes, museos y telecomunicaciones) y ocho idiomas (inglés, árabe, chino, alemán, francés, ruso, español y turco). En la competencia SemEval 2016 se propuso la evaluación de métodos para la extracción de aspectos a nivel de oración y a nivel de documento. Los conjuntos de datos de hoteles, restaurantes y laptops fueron anotados con el mismo esquema de anotación de SemEval 2015. Este conjunto de datos, al igual que los ofrecidos en SemEval 2014 y 2015, está disponible en formato XML a través de la plataforma META-SHARE. Los trabajos analizados en este SLR han empleado el conjunto de datos para el idioma inglés. La Tabla 6 muestra la cantidad de opiniones por dominios para el subconjunto de los datos en idioma inglés, donde SB1 corresponde a los textos disponibles para extraer aspectos a nivel de oración y SB2 corresponde a los textos disponibles para determinar los aspectos a nivel de documento.

Tabla 6. Cantidad de opiniones por dominios en el conjunto de datos de SemEval 2015 para el idioma inglés.

Dominio	**Entrenamiento**		**Prueba**		**Total**
	SB1	**SB2**	**SB1**	**SB2**	
Restaurante	2000	1950	676	676	5302
Laptop	2500	2375	808	808	6491

El conjunto de datos[5] obtenido a partir de TripAdvisor[6] [93] es usado por [109], [110] e incluye 174615 opiniones de 1768 hoteles, donde se abordan cinco aspectos: precio, habitación, ubicación, limpieza y servicio. A cada opinión se le asigna una puntuación en el rango de 1 a 5 estrellas.

Varios autores han usado los datos obtenidos del sitio web de Amazon[7] para evaluar sus propuestas. En [79] se utilizaron 12700 opiniones de teléfonos inteligentes publicadas en el sitio web de Amazon. En cada oración los términos fueron etiquetados con cinco aspectos predefinidos (batería, pantalla, cámara, altavoz, velocidad de ejecución). Las oraciones que tienen al menos un aspecto fueron etiquetadas con la polaridad (positiva o negativa). En el trabajo publicado en [83] se usa otro conjunto de datos[8] proveniente del sitio web de Amazon. Este conjunto contiene más de 340000

[5] http://times.cs.uiuc.edu/~wang296/Data/
[6] http://www.tripadvisor.com
[7] http://www.amazon.com/
[8] http://www.cs.jhu.edu/~mdredze/datasets/sentiment/

opiniones de 22 tipos de productos diferentes los que están etiquetados con la polaridad positiva o negativa. En [83] se empleó un subconjunto de esta colección que solo aborda cuatro dominios: libros, dvd, productos electrodomésticos y de cocina. Para cada dominio se tomaron 1000 opiniones positivas y 1000 negativas. En [111] se usa otro conjunto de datos[9] procedente del sitio web de Amazon y que fue construido y evaluado en [108]. Este conjunto de datos está asociado al dominio de opiniones sobre dispositivos para mp3 donde cada opinión tiene una escala de 1 a 5 estrellas (puntuación). En general, en los trabajos que han empleado datos de Amazon los datos han sido construidos durante el proceso de investigación o se han empleado conjuntos de datos públicos.

Otro conjunto de datos usado por varios investigadores [91], [93], [111] es el propuesto en la competición Yelp[10]. Este conjunto de datos consta de 5200000 opiniones sobre diferentes negocios y se encuentra disponible de forma pública. Las opiniones tienen una escala de 1 a 5 estrellas (puntuación) que indican la polaridad y otros criterios como útil, gracioso o interesante, con un valor numérico según los puntos o votos recibidos. Los datos se encuentran en formato JSON y SQL. Los métodos que han empleado este conjunto han trabajado solo con una pequeña parte de él. Por ejemplo, en [112] se toman 1598 opiniones sobre restaurants y 1335 sobre laptops. En [111] se usan 20000 opiniones sobre restaurantes. La selección de un conjunto pequeño (opiniones sobre restaurantes o laptops) [26], [111], [113] se debe a que se entrena con este conjunto de datos y se compara con otros conjuntos como los de SemEval 2014, SemEval 2016 o de Amazon.

Una característica presente en varios trabajos es que no realizan el entrenamiento y la evaluación con un único conjunto de datos [26], [80], [87], [92], [114], sino que entrenan con un conjunto (por ejemplo: Yelp, Amazon, SemEval 2014, 2015 y 2016) y evalúan con otro conjunto que en la mayoría de los casos es del mismo dominio del entrenamiento. El conjunto de datos más empleado por los investigadores es el de SemEval 2014. La selección se debe a que los datos están etiquetados con bastante granularidad (polaridad a nivel de aspecto y oración, categorías de aspectos). Características similares tiene los conjuntos de datos de SemEval 2015 y 2016, ya que se tomaron a partir del propuesto en SemEval 2014. La mayoría de los trabajos realizan el entrenamiento y evaluación sobre opiniones en los dominios de restaurantes y laptops. Los trabajos revisados no usan el mismo conjunto de datos para evaluar sus propuestas, lo que dificulta establecer comparaciones. El empleo de un mismo conjunto de datos permitiría tener un criterio más certero de la efectividad de cada propuesta.

[9] http://sifaka.cs.uiuc.edu/~wang296/Data/index.html
[10] https://www.yelp.com/dataset/challenge

RQ4: ¿Cuáles son las formas de representación textual más empleadas al extraer aspectos utilizando métodos de aprendizaje profundo?

La forma de representación textual está asociada a la posible organización de la información del texto no estructurado. Los textos generalmente se conforman por párrafos, oraciones y palabras. Una correcta organización de la información no estructurada es necesaria para lograr un efectivo entrenamiento de los métodos de aprendizaje profundo y consecuentemente realizar una eficaz extracción de aspectos. El aprendizaje profundo se basa esencialmente en el trabajo con redes neuronales. Uno de los retos más importantes al utilizar redes neuronales es lograr una forma de representación correcta para los datos de entrada a la red. Los conjuntos de entrenamiento de estas redes están formados por documentos u oraciones que contienen palabras.

Word Embeddings [45] se definió para lograr un mejor entrenamiento de las redes neuronales que se utilizan en NLP. Éste es el nombre de un conjunto de lenguajes de modelado y técnicas de aprendizaje dónde las palabras o frases del vocabulario se vinculan a vectores de números reales. Word Embeddings conceptualmente transforma un espacio con una dimensión por cada palabra a un espacio vectorial continuo con menos dimensiones [45]. Estos vectores se pueden crear a partir de un conjunto de palabras con la herramienta word2vec[11].

De las propuestas analizadas, el 93% emplea como forma de representación el Word Embeddings, como se muestra en la Figura 10. Éste necesita de grandes volúmenes de información para la creación de los vectores asociados a las palabras. Siguiendo la forma de representación Word Embeddings se han creado varios modelos, como se muestra en la Figura 11. Uno de los primeros modelos propuestos en la literatura fue Senna [115]. Para este modelo se define un vector de 50 dimensiones y los conjuntos de datos pueden ser entrenados a partir de las herramientas creadas por los autores[12]. En [46] se propone el modelo Glove[13], que se diferencia del anterior porque trata de capturar las estadísticas globales y al mismo tiempo las relaciones en el contexto donde aparecen las palabras. Con este modelo se pueden obtener vectores pre-entrenados con grandes conjuntos de información, como el existente en Wikipedia[14] en inglés. Los modelos Skip-gram y bolsa de palabras contextual (*Contextual Bag-of-Words*; CBOW) fueron propuestos por [45]. El primero predice, dada una palabra, las palabras del contexto o ventana. El objetivo de CBOW es predecir una palabra si se conoce el contexto o ventana de palabras.

[12] https://ronan.collobert.com/senna/
[13] https://nlp.stanford.edu/projects/glove/.
[14] https://en.wikipedia.org/wiki/Main_Page

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

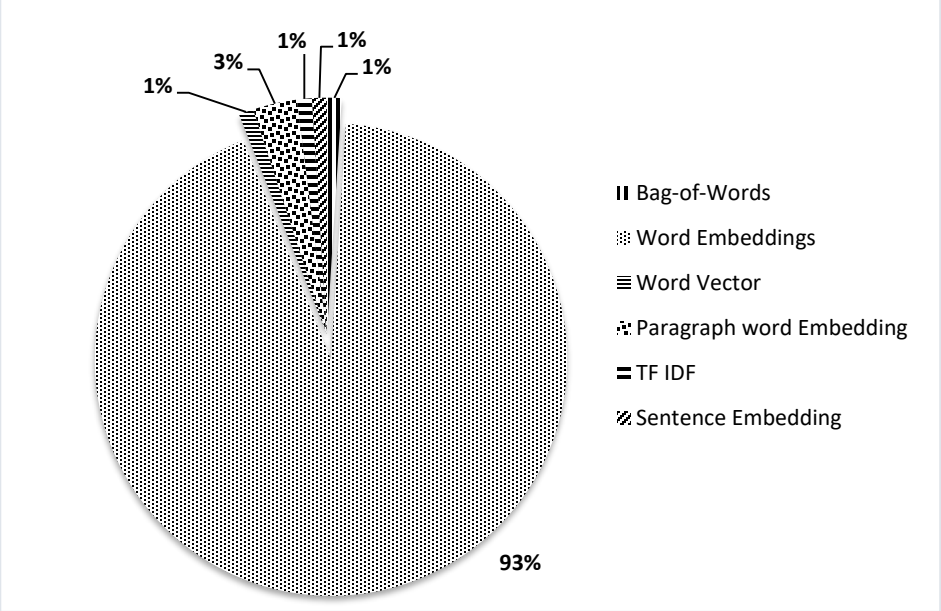

Figura 10. Formas de representación textual empleadas en las publicaciones acerca de la extracción de aspectos utilizando aprendizaje profundo.

En [45] se propuso e hizo público un conjunto de vectores pre-entrenados, a partir de un conjunto de datos de noticias procedentes del sitio Google News[15] (100 mil millones de palabras). El modelo contiene un vector de dimensión 300 para 3 millones de palabras o frases. Este conjunto de vectores fue empleado por siete de los trabajos analizados en este artículo de revisión que emplean el Word Embeddings en la extracción de aspectos utilizando métodos del aprendizaje profundo.

En varios trabajos [26], [87], [116] se usan conjuntos de datos (como el propuesto en Yelp[16]o Amazon[17]) para el entrenamiento del vector asociado a las palabras usando la herramienta word2vec. En la Figura 11 se muestran los resultados de la cantidad de veces en que son usados estos vectores. Auto CBOW y Auto Skip-Gram representan la cantidad de veces en que las propuestas realizan el entrenamiento de un Word Embeddings a partir de un conjunto de datos de entrenamiento y la herramienta word2vec. Fasttext[18], empleado en [117], [118] con muy buenos resultados, es un modelo de Word Embedding propuesto por [119] y que posee un modelo pre-entrenado. Este conjunto de vectores

[15] https://news.google.com
[16]https://www.yelp.com/datasetchallenge/
[17] http://www.cs.jhu.edu/~ mdredze/datasets/sentiment/
[18] https://fasttext.cc

pre-entrenados y la herramienta para el entrenamiento del corpus muestra mejores resultados que word2vec y Glove en términos de velocidad, escalabilidad y efectividad [98]. Se deben hacer evaluaciones en la tarea ABSA de estos datos con respecto a otras propuestas como word2vec, Senna, Sentiment WE, Glove.

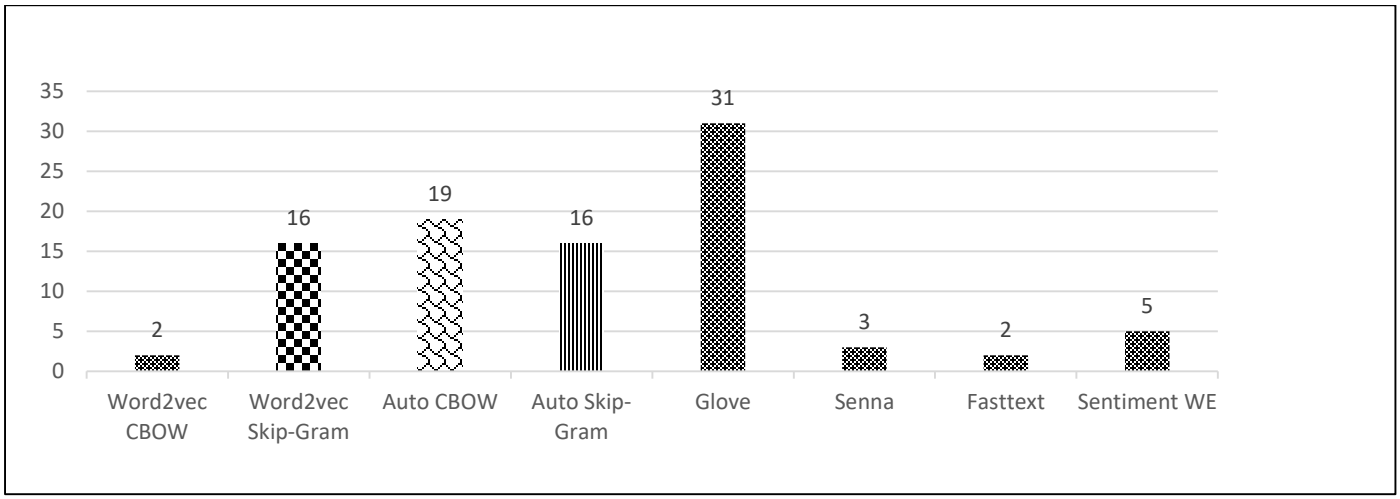

Figura 11. Modelos de Word Embeddings usados en algunos trabajos analizados.

Tabla 7. Características de los conjuntos de vectores Word Embeddings pre-entrenados.

Word Embeddings pre-entrenados	**Referencia**	**Dimensión**	**Tamaño del vocabulario**
Senna/Wikipedia	[115]	50	130000
word2vec CBOW/Google News	[45]	300	3000000
word2vec CBOW/Amazon	[74]	50, 300	1000000
$SSWE_h$, $SSWE_r$, $SSWE_u$	[120]	50	137000
GloVe/Wikipedia 2014 + Gigaword 5	[46]	50,100, 200, 300	1200000
Fasttext/English	[119]	Diferente (más de 82 idiomas)	Diferente (más de 82 idiomas)

Sentiment WE, conocido como el Word Embeddings específico de sentimientos (*Sentiment-Specific Word Embedding*[19]; SSWE) [120], representa un conjunto de Word Embeddings entrenados con tweets de sentimientos. Este conjunto de vectores de Word Embeddings tiene dimensión 50 y está dividido en $SSWE_h$, $SSWE_r$ y $SSWE_u$. Cada uno de estos conjuntos ha sido entrenado con diferentes algoritmos y solo usa oraciones de sentimientos. $SSWE_u$ tiene en

[19]http://ir.hit.edu.cn/˜dytang

cuenta simultáneamente la oración de sentimiento y el contexto donde ocurren las palabras. En las propuestas presentadas en [121], [122] se comparan los resultados obtenidos con estos conjuntos y los vectores pre-entrenados de Glove, pero no se obtienen mejores resultados. La Tabla 7 ofrece detalles de los Word Embeddings pre-entrenados.

Para la extracción de aspectos con métodos de aprendizaje profundo, en [113] se realiza una comparación entre los resultados con diferentes modelos: Skip-gram, CBOW y Glove. Durante la evaluación de esta propuesta los mejores resultados se alcanzaron al usar el modelo Glove. En la propuesta publicada en [114], se realiza el entrenamiento usando Word Embeddings de 200 dimensiones con word2vec y el conjunto de datos de Yelp en el dominio de opiniones de restaurantes. Además, emplearon para el dominio de laptop un Word Embeddings de vectores pre-entrenados con el modelo Glove.

El concepto de vector de valores reales asociados a palabras puede ser extendido a oraciones o párrafos, a partir del conjunto de datos de entrenamiento. En [85] se realiza el entrenamiento del vector de valores reales en función de las oraciones presentes en el conjunto de datos y en el trabajo publicado en [110] se usan los párrafos. Estas formas de representación de la información repercutieron negativamente en la calidad de la extracción de aspectos.

En [83] se emplea una bolsa de palabras (*bag-of-words*) y se obtiene un vector binario que codifica la presencia/ausencia de unigramas y bigramas. En [84] se tiene un vector de palabras y se calcula la frecuencia de los términos (*Term Frecuency*; TF) para los sustantivos en el conjunto de datos de entrenamiento y se calcula la frecuencia inversa del documento (*Inverse Document Frecuency*; IDF) en un conjunto de datos de *n*-gramas de Google[20]. Las propuestas que utilizan estas formas de representación del conocimiento no necesitan grandes conjuntos de datos para su entrenamiento y no utilizan una representación vectorial de grandes dimensiones; sin embargo, pierden la riqueza semántica que posee el Word Embeddings. Por esta razón, no superan los resultados alcanzados por aquellas propuestas que utilizan Word Embeddings.

Algunos autores evalúan la posibilidad de agregar al vector del modelo Word Embeddings más entradas asociadas a las características del contexto donde aparecen las palabras [80], [123]. En [54] se emplea un vector de Word Embeddings de 300 dimensiones. A este vector se le añaden seis entradas asociadas a seis tipos de etiquetas morfológicas que puede tener la palabra (sustantivo, verbo, adjetivo, adverbio, preposición, conjunción) y se codifican estas entradas de forma binaria según las características morfológicas de la palabra. En esta propuesta, el Word Embeddings con las entradas de las etiquetas morfológicas mejoran los resultados de evaluación del método. En [91] se utiliza el vector pre-entrenado Glove, donde se le añade a cada palabra una entrada con la distancia relativa al aspecto presente en la oración y otra

[20] http://books.google.com/ngrams/datasets

entrada para la etiqueta morfológica de la palabra con el objetivo que ayuden en el proceso de aprendizaje. Los resultados obtenidos por el método propuesto en este trabajo superan al resto de los métodos con los que es comparado. Lo antes expuesto reafirma que Word Embeddings es la forma de representación del conocimiento más empleada en los artículos analizados que realizan la extracción de aspectos con métodos de aprendizaje profundo. En estos artículos son usados varios modelos como: Senna, Glove, Skip-gram y CBOW. La forma en que son entrenados estos modelos y el uso de los vectores pre-entrenados, como el que usa información de Google News y los de Glove, influyen en los resultados finales. Para poder seleccionar el mejor modelo, se debe investigar más sobre la influencia de estos modelos en los resultados.

RQ5: ¿Cuáles son las medidas y los dominios del conocimiento que más se utilizan para evaluar el desempeño de los métodos de aprendizaje profundo en la extracción de aspectos?

Los métodos analizados han sido validados mediante la aplicación de diversas medidas de calidad. Las medidas para la evaluación de los resultados de las diferentes propuestas se concentran en el uso de la Exactitud (*Accuracy*), Exhaustividad (*Recall*) y Micro F1; no obstante, otras medidas también han sido empleadas, ellas son: la media del Error Cuadrático Medio (*Root Mean Square Error*), la Correlación local del aspecto en la opinión (*Aspect correlation inside reviews*) y la correlación global de aspectos en todas las opiniones (*Aspect correlation across all reviews*), como se muestra en la Figura 12.

Las medidas *Accuracy*, *Precision*, *Recall* y *Micro F1* son muy usadas para evaluar la clasificación en problemas de NLP. Éstas cuantifican la calidad de la extracción de aspectos teniendo en cuenta la cantidad de aspectos extraídos de forma correcta con respecto a los aspectos de referencia. La Perplejidad (*Perplexity*) es otra medida de validación empleada para analizar la calidad en la identificación de aspectos [84], [111]. Esta medida cuantifica, para cada documento, la cantidad de aspectos encontrados respecto a la cantidad total de palabras del documento. En [85], [111] se emplea la medida Coherencia de Tópicos (*Topic Coherence*), propuesta en [124]. Ésta es una medida de la calidad de los aspectos basada en la coocurrencia de palabras. Según esta medida, el método de aprendizaje profundo propuesto en [85] tiene mejores resultados que LDA al extraer aspectos. En [109], [110] se emplean las medidas *Aspect correlation inside reviews* y *Aspect correlation across all reviews* para evaluar la calidad de la extracción de aspectos. La primera medida intenta medir cuán bien el método puede mantener el orden relativo de los aspectos dentro de la opinión (a nivel de oración). La segunda indica si los valores obtenidos y los valores del conjunto de prueba para un aspecto dado daría una clasificación similar en todas las opiniones donde aparece este aspecto.

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

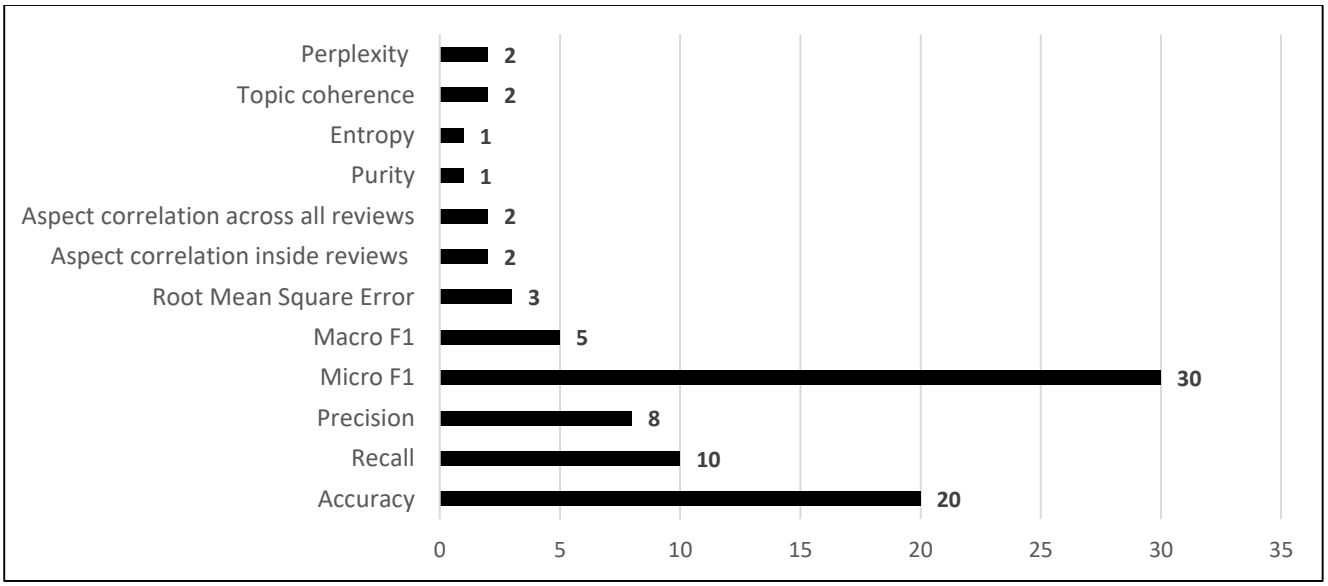

Figura 12. Medidas de evaluación usadas para calcular la calidad de las propuestas que aplican aprendizaje profundo para la extracción de aspectos.

Muchas veces se dificulta la comparación entre las propuestas existentes, ya sea porque utilizan diferentes medidas de validación o diferentes colecciones (que en algunos casos no están disponibles o se utiliza un subconjunto de aquellas disponibles); por ejemplo, en las publicaciones [25], [72], [84], [90], [101], [125] no se realiza una comparación con otras investigaciones que han reportado buenos resultados empleando CRF o LDA [34], [126], [127] .

Algunas propuestas solo miden sus resultados a través del *Accuracy* [88], [91], [101], [112], [128]. Estos resultados no permiten determinar correctamente la calidad del método. Esto se debe a que solamente se tienen en cuenta la relación entre los datos correctos y el total de aspectos extraídos por el método y se excluye la información del conjunto de pruebas. El uso de las medidas *Accuracy*, *Recall* y *Micro F1* por varios autores no es aislado (se muestran los tres resultados para poder demostrar la validez de los métodos). *Micro F1* es una medida que combina los resultados del *Accuracy* y el *Recall* [129]. Por esta razón es seleccionada para identificar cuáles de los trabajos analizados arrojan los mejores resultados.

Para el caso de los métodos que extraen aspectos asociados a categorías (*Aspect-Category Sentiment Analysis*; ACSA), el mejor resultado de *Micro-F1* es el obtenido en [79], alcanzándose un 93.63% al emplear CNN a un conjunto de datos de opiniones de Amazon sobre teléfonos inteligentes. En esta propuesta se probó contra un SVM usando validación cruzada. Una valoración más completa de este método debería incluir su comparación respecto a métodos que utilicen

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

CRF o LDA. El segundo mejor resultado en la subtarea ACSA es el obtenido en [130] donde se obtiene un 88.91% de *Micro-F1* al entrenarse con el conjunto de datos propuesto en la competición SemEval 2014.

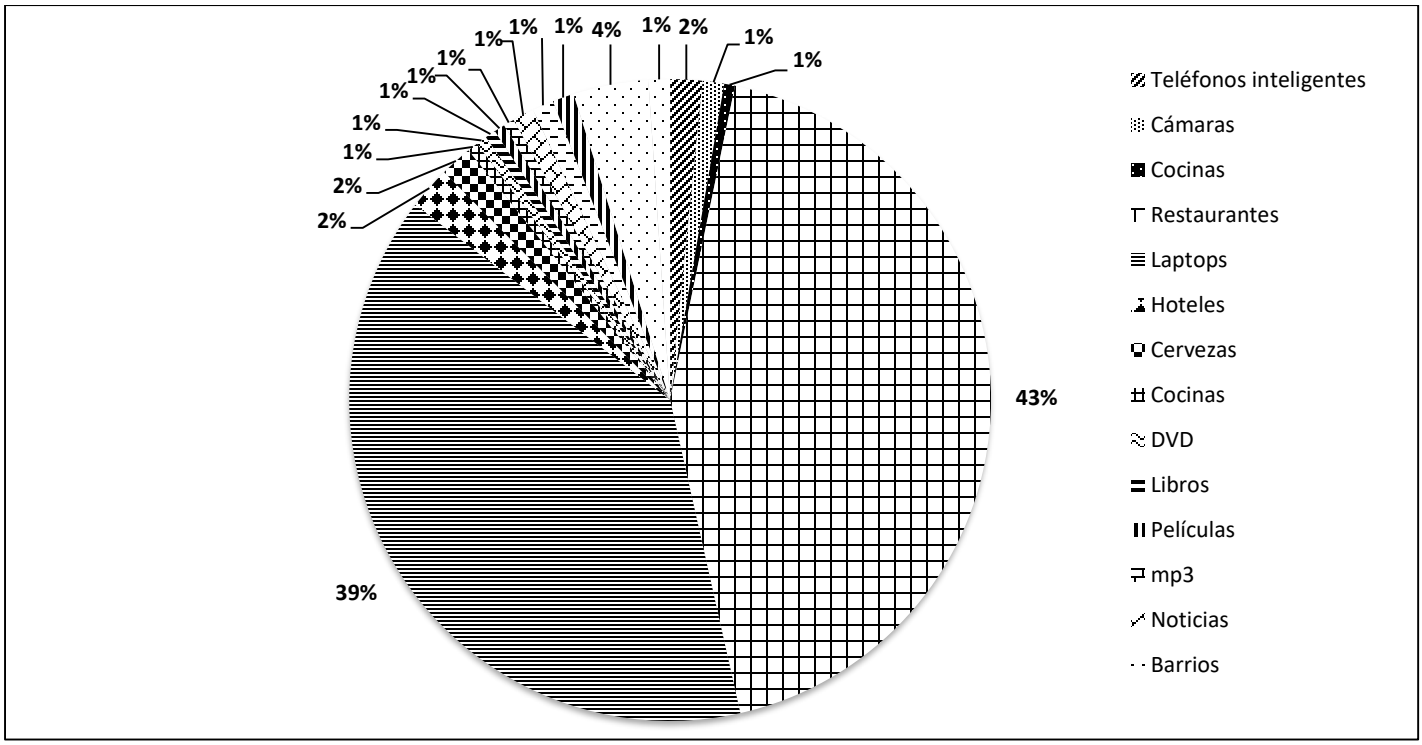

Figura 13. Dominios del conocimiento empleados en las evaluaciones.

Para la extracción de aspectos el mejor resultado teniendo en cuenta la medida de evaluación *Micro-F1* es el reportado en [80], obteniéndose un 82% para el dominio Laptop y un 87.7% para restaurantes. El método propuesto en este trabajo es CNN, aunque se auxilia de reglas lingüísticas propuestas por los autores y que tienen influencia en los resultados finales. La otra propuesta que tiene buenos resultados es [131], donde se obtiene un 85.29% de *Micro-F1* al ser evaluado en el conjunto de datos SemEval 2014 con opiniones sobre restaurantes. En este mismo conjunto de datos para opiniones de laptops se alcanza un 77.80%. En este trabajo presentado en [131] se emplea un GRU con mecanismo de atención y se extraen a la misma vez aspectos y términos de opinión. El método es evaluado con otras propuestas que usan técnicas de aprendizaje profundo y CRF. En este trabajo se muestra como otros métodos que combinan aprendizaje profundo y características lingüísticas obtienen resultados similares. Este método en dominios diferentes (laptops, restaurantes) obtiene resultados diferentes, y en algunos casos, con una marcada diferencia. Esto es un inconveniente de los métodos propuestos porque en algunas ocasiones son dependientes del dominio de aplicación y, por tanto, se requiere el

Revista Cubana de Ciencias Informáticas
Vol. x, No. x, Mes x-y, 2013
ISSN: 2227-1899 | RNPS: 2301
Pág. 105-145
http://rcci.uci.cu

desarrollo de sistemas específicos para cada dominio. Una gran parte de los métodos propuestos realiza la evaluación solamente con opiniones sobre dominios de restaurantes y laptops. El uso de estos dominios está dado por la existencia de los conjuntos de datos de las competiciones SemEval 2014, 2015 y 2016 y las referencias de los mejores resultados en estas competiciones para poder comparar. Esto pudiera ocultar errores para el ABSA en otros dominios del conocimiento, incluso existiendo la disponibilidad de conjuntos de datos. La propuesta presentada en [97] asume el reto de aprender en múltiples dominios empleando técnicas de Aprendizaje Profundo y enfrentando el problema del olvido catastrófico. El olvido catastrófico ocurre cuando se deben aprender secuencialmente varias tareas empleando redes neuronales [132]. En la propuesta referenciada en [97] se obtiene un *accuracy* de 72.73% y un *Micro F1* de 67.92%. Aunque estos resultados son bajos, este trabajo indica la posibilidad de lograr resultados en el reto de obtener buenos resultados en ABSA para varios dominios empleando modelos del Aprendizaje Profundo.

En la Figura 13 el dominio restaurantes es el más utilizado al evaluar las propuestas existentes para la extracción de aspectos. Es interesante destacar que en este dominio es frecuente encontrar términos de opinión o aspectos sobre distintos servicios y productos que se pueden ofrecer en un restaurante. Este es un dominio diverso por los diferentes temas que pueden ser abordados por los usuarios.

Resultados y discusión

El análisis de los 89 artículos sobre el uso del aprendizaje profundo para la extracción de aspectos publicados desde enero de 2011 hasta febrero de 2019 arrojó los siguientes resultados:

- LSTM es la técnica del aprendizaje profundo más empleada en la extracción de aspectos. La selección de esta técnica por parte de los investigadores se justifica por la variedad de problemas del NLP resueltos aplicando LSTM y la naturaleza secuencial de los datos en la tarea ABSA.
- Word Embeddings con vectores asociados a palabras es la forma de representación más empleada para la extracción de aspectos aplicando aprendizaje profundo. Dentro de éste, el modelo más usado es skip-gram con un conjunto de vectores pre-entrenados con la información de Google News. No obstante, algunos investigadores emplean exitosamente otros conjuntos de datos para el entrenamiento inicial del Word Embeddings. Esta forma de representación es muy útil para datos de entradas de las redes neuronales de los algoritmos del aprendizaje profundo debido a que representan un vector de números reales. Los grandes conjuntos de datos con los que se crean los vectores permiten cubrir gran cantidad de ejemplos y representar la relación semántica entre palabras, información útil en el proceso de extracción de aspectos.

- El conjunto de datos más empleado por los investigadores es el de SemEval 2014. La selección de éste se debe a que los datos están etiquetados con bastante granularidad (polaridad a nivel de aspecto y oración, categorías de aspectos) y la cantidad de información disponible para el proceso de entrenamiento y prueba. En el aprendizaje y la evaluación también se pueden usar otros conjuntos de datos como los de Yelp y Amazon. La mayoría de los trabajos realizan el entrenamiento y la evaluación sobre opiniones en los dominios de restaurantes y laptops. Esto sucede porque son los dominios más frecuentes y con más ejemplos en los conjuntos de datos.
- Varias propuestas no logran mejores resultados que otras máquinas de aprendizaje como SVM, LDA o CRF. Otros trabajos analizados no evalúan los métodos propuestos con estas máquinas de aprendizaje.
- Algunas propuestas evalúan su calidad solamente con el *Accuracy*. Se debe evitar usar solamente una medida de calidad porque no se tiene un criterio acertado acerca de la calidad de los métodos propuestos.
- El uso de reglas lingüísticas combinadas con técnicas de aprendizaje profundo logra mejorar los resultados de los métodos al ser evaluados con otras propuestas como CRF o LDA.
- En el futuro los autores deben evaluar el impacto del uso de recursos externos como lexicones, redes de conceptos y técnicas de aprendizaje profundo no supervisadas.

Es necesario continuar las investigaciones en el uso de métodos de aprendizaje profundo para la tarea ABSA. Los principales resultados de esta investigación sugieren que se debe profundizar en la identificación del mejor modelo de Word Embeddings, evaluar dominios diferentes al de restaurantes y laptops, y realizar comparaciones con propuestas entrenadas con CRF o LDA.

Conclusiones

En esta investigación se logró agrupar y evaluar varios de trabajos que no se analizaron en 27 artículos de revisión sobre el análisis de sentimiento o la tarea ABSA. De estos trabajos se han logrado determinar las formas de representación, modelos, resultados y conjuntos de datos empleados. Sin embargo, la cantidad de trabajos es insuficiente. Los resultados sobre dominios del conocimiento (restaurante, hoteles, laptop, entre otros) evidencia la necesidad de definir propuestas que pueda tener buenos resultados en múltiples dominios del conocimiento. La mayoría de los métodos que usan técnicas de aprendizaje profundo siguen un enfoque supervisado, por lo que requieren partir de colecciones previamente clasificadas. Una línea de investigación pudiera estar dirigida a seguir profundizando en el desarrollo de propuestas no supervisadas o híbridas para la extracción de aspectos. Las nuevas propuestas de métodos de aprendizaje profundo deben ser evaluadas contra otras máquinas de aprendizaje, para poder estimar correctamente su aporte. Se debe

continuar investigando en la identificación y selección de los mejores modelos para la representación textual. Futuros análisis de la literatura se deben orientar al estudio, consolidación, clasificación y crítica en general de los métodos de aprendizaje profundo en la tarea ABSA para múltiples dominios. Es necesaria la investigación y creación de propuestas que usen recursos externos como lexicones y redes de conceptos combinado con técnicas de aprendizaje profundo (principalmente no supervisadas) para la tarea ABSA. Las competiciones SemEval 2017 y SemEval 2018 no dedicaron tareas a la evaluación del ABSA pero otras como ESWC 2017 y 2018 [133] mantienen activa las propuestas de nuevos métodos de extracción de aspectos y análisis de sentimientos. De conjunto a nuevas propuestas para ABSA, uno de los campos de investigación que ha tomado auge para el análisis de opiniones en todos sus niveles es la detección de opiniones falsas (*fake opinions*).

Referencias

[1] M. Hu y B. Liu, «Mining and summarizing customer reviews», en *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, pp. 168–177.
[2] T. Nasukawa y J. Yi, «Sentiment analysis: capturing favorability using natural language processing», en *Proceedings of the 2nd International Conference on Knowledge Capture*, 2003, pp. 70-77.
[3] S. Das y M. Chen, «Yahoo! for Amazon: extracting market sentiment from stock message boards», en *Proceedings of the Asia Pacific Finance Association Annual Conference (APFA)*, 2001, vol. 35, p. 43.
[4] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press, 2015.
[5] B. Pang y L. Lee, *Opinion mining and sentiment analysis*, vol. 2. 2008.
[6] B. Liu, *Sentiment analysis and opinion mining*, 1.ª ed., vol. 5. Morgan & Claypool, 2012.
[7] S. M. Jiménez Zafra, E. Martínez Cámara, M. T. Martín Valdivia, y M. D. Molina González, «Tratamiento de la negación en el análisis de opiniones en español», *Procesamiento del Lenguaje Natural, Revista nº 54*, pp. 37–44, 2015.
[8] S. Sarawagi, «Information extraction», *Foundations and Trends in Databases*, vol. 1, n.º 3, pp. 261–377, 2008.
[9] N. Indurkhya y F. J. Damerau, *Handbook of natural language processing*, CRC Machin., vol. 2. CRC Press, 2010.
[10] S. Gottipati y J. Jiang, «Linking entities to a knowledge base with query expansion», en *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2011, pp. 804-813.
[11] M. Dredze, P. McNamee, D. Rao, A. Gerber, y T. Finin, «Entity disambiguation for knowledge base population», en *Proceedings of the 23rd International Conference on Computational Linguistics*, 2010, pp. 277-285.
[12] X. Ding, B. Liu, y P. S. Yu, «A holistic lexicon-based approach to opinion mining», en *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 2008, pp. 231–240.
[13] A. Valitutti, C. Strapparava, y O. Stock, «Developing affective lexical resources», *PsychNology Journal*, vol. 2, n.º 1, pp. 61–83, 2004.

[14] F. Li *et al.*, «Structure-aware review mining and summarization», en *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, 2010, pp. 653–661.
[15] Z. Chen y B. Liu, «Lifelong machine learning», *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 10, n.º 3, pp. 1-145, 2016.
[16] B. Liu, «Sentiment analysis and subjectivity», *Handbook of Natural Language Processing*, vol. 2, pp. 627–666, 2010.
[17] L. Zhang y B. Liu, «Aspect and entity extraction for opinion mining», en *Data Mining and Knowledge Discovery for Big Data*, Springer, 2014, pp. 1-40.
[18] Q. Su *et al.*, «Hidden sentiment association in chinese web opinion mining», en *Proceedings of the 17th International Conference on World Wide Web*, 2008, pp. 959–968.
[19] Z. Hai, K. Chang, y J. Kim, «Implicit feature identification via co-occurrence association rule mining», en *Computational Linguistics and Intelligent Text Processing*, Springer, 2011, pp. 393–404.
[20] G. Fei, B. Liu, M. Hsu, M. Castellanos, y R. Ghosh, «A dictionary-based approach to identifying aspects implied by adjectives for opinion mining», en *24th International Conference on Computational Linguistics*, 2012, p. 309.
[21] M. Pontiki *et al.*, «SemEval-2016 Task 5: Aspect Based Sentiment Analysis», en *Proceedings of the 10th international workshop on semantic evaluation (SemEval 2016)*, San Diego, California, USA, 2016, pp. 19-30.
[22] I. La Vie, «Taming the hashtag: universal sentiment, SPEQ-ing the truth, and structured opinion in social media», Iowa State University, 2015.
[23] E. Phillips, «The universal voting markup language (uvml)», *age*, vol. 1, p. 3DIGIT, 2013.
[24] S. Aoki y O. Uchida, «A method for automatically generating the emotional vectors of emoticons using weblog articles», en *Proceedings of the 10th WSEAS International Conference on Applied Computer and Applied Computational Science*, 2011, pp. 132-136.
[25] T. H. Nguyen y K. Shirai, «PhraseRNN: phrase recursive neural network for aspect-based sentiment analysis», en *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Lisboa, Portugal, 2015, pp. 2509-2514.
[26] H. Ye, Z. Yan, Z. Luo, y W. Chao, «Dependency-tree based convolutional neural networks for aspect term extraction», en *Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Jeju, South Korea, 2017, vol. 10235, pp. 350-362.
[27] D. Tang, B. Qin, y T. Liu, «Deep learning for sentiment analysis: successful approaches and future challenges», *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 5, n.º 6, pp. 292-303, 2015.
[28] A.-M. Popescu, B. Nguyen, y O. Etzioni, «OPINE: extracting product features and opinions from reviews», en *Proceedings of HLT/EMNLP on Interactive Demonstrations*, 2005, pp. 32–33.
[29] S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. A. Reis, y J. Reynar, «Building a sentiment summarizer for local service reviews», en *WWW Workshop on NLP in the Information Explosion Era*, 2008, vol. 14, pp. 339-348.
[30] M. Hu *et al.*, «Opinion extraction, summarization and tracking in news and blog corpora», en *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, 2010, vol. 100107, pp. 261–377.

[31] C. Long, J. Zhang, y X. Zhut, «A review selection approach for accurate feature rating estimation», en *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, 2010, pp. 766–774.
[32] S. Moghaddam y M. Ester, «Opinion digger: an unsupervised opinion miner from unstructured product reviews», en *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 2010, pp. 1825–1828.
[33] L. Zhuang, F. Jing, y X.-Y. Zhu, «Movie review mining and summarization», en *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, 2006, pp. 43–50.
[34] G. Qiu, B. Liu, J. Bu, y C. Chen, «Opinion word expansion and target extraction through double propagation», *Computational Linguistics*, vol. 37, n.º 1, pp. 9–27, 2011.
[35] Z. Zhai, B. Liu, H. Xu, y P. Jia, «Clustering product features for opinion mining», en *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 2011, pp. 347-354.
[36] Y. Choi, E. Breck, y C. Cardie, «Joint extraction of entities and relations for opinion recognition», en *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 431–439.
[37] B. Yang y C. Cardie, «Extracting opinion expressions with semi-markov conditional random fields», en *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 1335-1345.
[38] T. Hofmann, «Probabilistic latent semantic indexing», en *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1999, pp. 50-57.
[39] D. M. Blei, A. Y. Ng, y M. I. Jordan, «Latent dirichlet allocation», *Journal of Machine Learning Research*, vol. 3, n.º Jan, pp. 993-1022, 2003.
[40] F. Li, M. Huang, y X. Zhu, «Sentiment analysis with global topics and local dependency», en *Association for the Advancement of Artificial Intelligence*, 2010, vol. 10, pp. 1371–1376.
[41] I. Titov y R. McDonald, «Modeling online reviews with multi-grain topic models», en *Proceedings of the 17th International Conference on World Wide Web*, 2008, pp. 111–120.
[42] S. R. K. Branavan, H. Chen, J. Eisenstein, y R. Barzilay, «Learning document-level semantic properties from free-text annotations», *Journal of Artificial Intelligence Research*, pp. 569–603, 2009.
[43] L. Deng y D. Yu, «Deep learning: methods and applications», *Foundations and Trends® in Signal Processing*, vol. 7, n.º 3–4, pp. 197-387, 2014.
[44] D. Budgen y Pearl Brereton, «Guidelines for performing systematic literature reviews in software engineering», en *Proceedings of the 28th International Conference on Software Engineering*, Shanghai, China, 2006, pp. 1051-1052.
[45] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, y J. Dean, «Distributed representations of words and phrases and their compositionality», *Advances in Neural Information Processing Systems*, pp. 3111–3119, 2013.
[46] J. Pennington, R. Socher, y C. Manning, «Glove: Global vectors for word representation», en *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532-1543.
[47] S. Sun, C. Luo, y J. Chen, «A review of natural language processing techniques for opinion mining systems», *Information Fusion*, vol. 36, pp. 10-25, 2017.

[48] P. More y A. Ghotkar, «A study of different approaches to aspect-based opinion mining», *International Journal of Computer Applications*, vol. 145, n.º 6, pp. 11-15, 2016.
[49] E. Cambria, «Affective computing and sentiment analysis», *IEEE Intelligent Systems*, vol. 31, n.º 2, pp. 102-107, 2016.
[50] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, y M. Pantic, «A survey of multimodal sentiment analysis», *Image and Vision Computing*, vol. 65, pp. 3-14, 2017.
[51] A. Yousif, Z. Niu, J. K. Tarus, y A. Ahmad, «A survey on sentiment analysis of scientific citations», *Artificial Intelligence Review*, pp. 1-34, 2017.
[52] T. Al-Moslmi, N. Omar, S. Abdullah, y M. Albared, «Approaches to cross-domain sentiment analysis: a systematic literature review», *IEEE Access*, vol. 5, pp. 16173-16192, 2017.
[53] A. P. Kirilenko, S. O. Stepchenkova, H. Kim, y X. Li, «Automated sentiment analysis in tourism: comparison of approaches», *Journal of Travel Research*, p. 47287517729757, 2017.
[54] A. Yadollahi, A. G. Shahraki, y O. R. Zaiane, «Current state of text sentiment analysis from opinion to emotion mining», *ACM Computing Surveys (CSUR)*, vol. 50, n.º 2, p. 25, 2017.
[55] R. S. Ramya, K. R. Venugopal, S. S. Iyengar, y L. M. Patnaik, «Feature extraction and duplicate detection for text Mining: a survey», *Global Journal of Computer Science and Technology*, vol. 16, n.º 5, pp. 1-20, 2017.
[56] Q. T. Ain, M. Ali, A. Riaz, y A. Noureen, «Sentiment analysis using deep learning techniques: a review», *International Journal of Advanced Computer Science and Applications*, vol. 8, n.º 6, pp. 424-433, 2017.
[57] L. Zhang, S. Wang, y B. Liu, «Deep learning for sentiment analysis: a survey», *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, p. 1253, 2018.
[58] Y. LeCun, Y. Bengio, y G. Hinton, «Deep learning», *Nature*, vol. 521, n.º 7553, pp. 436-444, 2015.
[59] G. E. Hinton y R. R. Salakhutdinov, «Reducing the dimensionality of data with neural networks», *Science*, vol. 313, n.º 5786, pp. 504-507, 2006.
[60] X. Glorot, A. Bordes, y Y. Bengio, «Deep sparse rectifier neural networks», en *Aistats*, 2011, vol. 15, p. 275.
[61] K. He, X. Zhang, S. Ren, y J. Sun, «Deep residual learning for image recognition», en *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
[62] I. Sutskever, J. Martens, G. E. Dahl, y G. E. Hinton, «On the importance of initialization and momentum in deep learning», *International Conference on Machine Learning, ICML (3)*, vol. 28, n.º 1139-1147, p. 5, 2013.
[63] S. Hochreiter y J. Schmidhuber, «Long short-term memory», *Neural Computation*, vol. 9, n.º 8, pp. 1735-1780, 1997.
[64] L. Deng, «A tutorial survey of architectures, algorithms, and applications for deep learning», *APSIPA Transactions on Signal and Information Processing*, vol. 3, 2014.
[65] L. Bottou, «Large-scale machine learning with stochastic gradient descent», en *Proceedings of COMPSTAT'2010*, Springer, 2010, pp. 177-186.
[66] D. Williams y G. Hinton, «Learning representations by back-propagating errors», *Nature*, vol. 323, n.º 6088, pp. 533-538, 1986.
[67] Y. LeCun, «Generalization and network design strategies», *Connectionism in perspective*, pp. 143-155, 1989.

[68] J. Chung, C. Gulcehre, K. Cho, y Y. Bengio, «Gated feedback recurrent neural networks», en *International Conference on Machine Learning*, 2015, pp. 2067-2075.
[69] H. Bourlard y Y. Kamp, «Auto-association by multilayer perceptrons and singular value decomposition», *Biological Cybernetics*, vol. 59, n.º 4, pp. 291-294, 1988.
[70] G. E. Hinton y R. S. Zemel, «Autoencoders, minimum description length and helmholtz free energy», *Advances in Neural Information Processing Systems*, pp. 3-10, 1994.
[71] P. Smolensky, «Information processing in dynamical systems: foundations of harmony theory», Colorado University at Boulder, Department of Computer Science, 1986.
[72] H. Wu, Y. Gu, S. Sun, y X. Gu, «Aspect-based opinion summarization with convolutional neural networks», en *International Joint Conference on Neural Networks (IJCNN), 2016*, 2016, pp. 3157-3163.
[73] H. H. Dohaiha, P. W. C. Prasad, A. Maag, y A. Alsadoon, «Deep learning for aspect-based sentiment analysis: a comparative review», *Expert Systems with Applications*, 2018.
[74] S. Joty, P. Liu, y H. M. Meng, «Fine-grained opinion mining with recurrent neural networks and word embeddings», en *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisboa, Portugal, 2015, pp. 1433-1443.
[75] B. Huang, Y. Ou, y K. M. Carley, «Aspect level sentiment classification with attention-over-attention neural networks», en *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 2018, pp. 197-206.
[76] J. Wang *et al.*, «Aspect sentiment classification with both word-level and clause-level attention networks», en *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, 2018, pp. 4439-4445.
[77] P. Chen, Z. Sun, L. Bing, y W. Yang, «Recurrent attention network on memory for aspect sentiment analysis», en *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 452-461.
[78] A. Chaudhuri y S. K. Ghosh, «Sentiment analysis of customer reviews using robust hierarchical bidirectional recurrent neural network», en *Artificial Intelligence Perspectives in Intelligent Systems*, Springer, 2016, pp. 249-261.
[79] X. Gu, Y. Gu, y H. Wu, «Cascaded convolutional neural networks for aspect-based opinion summary», *Neural Processing Letters*, vol. 46, pp. 581-594, 2017.
[80] S. Poria, E. Cambria, y A. Gelbukh, «Aspect extraction for opinion mining with a deep convolutional neural network», *Knowledge-Based Systems*, vol. 108, pp. 42-49, 2016.
[81] C. Sun, X. Wang, Y. Liu, B. Wang, y X. Wang, «Predicting polarities of tweets by composing word embeddings with long short-term memory», en *Proceeding of Association for Computational Linguistics (ACL).*, Beijing, China, 2015, pp. 1343-1353.
[82] J. Yuan, Y. Zhao, B. Qin, y T. Liu, «Local contexts are effective for neural aspect extraction», en *Proceedings of Chinese National Conference on Social Media Processing*, Beijing, China, 2017, pp. 244-255.
[83] X. Glorot, A. Bordes, y Y. Bengio, «Domain adaptation for large-scale sentiment classification: a deep learning approach», en *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, Bellevue, Washington, USA, 2011, pp. 513-520.

[84] L. Wang, K. Liu, Z. Cao, J. Zhao, y G. de Melo, «Sentiment-aspect extraction based on restricted boltzmann machines», en *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 2015, pp. 616-625.
[85] R. He, W. S. Lee, H. T. Ng, y D. Dahlmeier, «An unsupervised neural attention model for aspect extraction», en *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, 2017, vol. 1, pp. 388-397.
[86] S. Xiong, Y. Zhang, D. Ji, y Y. Lou, «Distance metric learning for aspect phrase grouping», en *Proceedings of the 2016 International Conference on Computational Linguistics (Coling)*, Osaka, Japon, 2016, pp. 2492–2502.
[87] J. Cheng, S. Zhao, J. Zhang, I. King, X. Zhang, y H. Wang, «Aspect-level sentiment classification with heat (hierarchical attention) network», en *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Singapore, Singapore, 2017, pp. 97-106.
[88] M. Huang, Y. Wang, X. Zhu, y L. Zhao, «Attention-based LSTM for aspect-level sentiment classification», en *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas, USA, 2016, pp. 606-615.
[89] D. Ying, J. Yu, y J. Jiang, «Recurrent neural networks with auxiliary labels for cross-domain opinion target extraction», en *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, San Francisco, USA, 2017, pp. 3436-3442.
[90] B.-D. Nguyen-Hoang, Q.-V. Ha, y M.-Q. Nghiem, «Aspect-based sentiment analysis using word embeddings restricted boltzmann machines», en *Proceedings of International Conference on Computational Social Networks*, 2016, vol. 9795, pp. 285-297.
[91] L. Xu, J. Lin, L. Wang, C. Yin, y J. Wang, «Deep convolutional neural network based approach for aspect-based sentiment analysis», *Advanced Science and Technology Letters*, vol. 143, pp. 199-204, 2017.
[92] W. Wang, S. J. Pan, D. Dahlmeier, y X. Xiao, «Recursive neural conditional random fields for aspect-based sentiment analysis», en *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, (EMNLP)*, Austin, Texas, USA, 2016, pp. 616-626.
[93] F. Xianghua, L. Guo, G. Yanyan, y W. Zhiqiang, «Multi-aspect sentiment analysis for chinese online social reviews based on topic modeling and hownet lexicon», *Knowledge-Based Systems*, vol. 37, pp. 186–195, 2013.
[94] B. Wang y W. Lu, «Learning latent opinions for aspect-level sentiment classification», 2018.
[95] L. Mai y B. Le, «Aspect-based sentiment analysis of vietnamese texts with deep learning», en *Asian Conference on Intelligent Information and Database Systems*, 2018, pp. 149-158.
[96] W. Wang, V. W. Zheng, H. Yu, y C. Miao, «A survey of zero-shot learning: settings, methods, and applications», *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, n.º 2, p. 13, 2019.
[97] S. Wang, G. Lv, S. Mazumder, G. Fei, y B. Liu, «Lifelong learning memory metworks for aspect sentiment classification», en *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 861-870.

[98] T. Young, D. Hazarika, S. Poria, y E. Cambria, «Recent trends in deep learning based natural language processing», *IEEE Computational Intelligence Magazine*, vol. 13, n.º 3, pp. 55-75, 2018.
[99] G. Ganu, N. Elhadad, y A. Marian, «Beyond the stars: improving rating predictions using review text content», en *WebDB*, 2009, vol. 9, pp. 1-6.
[100] J. Wiebe, T. Wilson, y C. Cardie, «Annotating expressions of opinions and emotions in language», *Language Resources and Evaluation*, vol. 39, n.º 2-3, pp. 165-210, 2005.
[101] D. Tang, B. Qin, y T. Liu, «Aspect level sentiment classification with deep memory network», en *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 214-224.
[102] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, y S. Manandhar, «Semeval-2014 task 4: aspect based sentiment analysis», en *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 2014, pp. 27–35.
[103] Y. Ma, H. Peng, y E. Cambria, «Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm», en *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, 2018, pp. 5876-5883.
[104] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, y S. Manandhar, «SemEval-2015 task 12: aspect based sentiment analysis», en *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Denver, Colorado, USA, 2015, pp. 486-495.
[105] S. Ruder, P. Ghaffari, y J. G. Breslin, «INSIGHT-1 at semeval-2016 task 5: deep learning for multilingual aspect-based sentiment analysis», en *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)*, 2016.
[106] Z. Toh y J. Su, «NLANGP at SemEval-2016 Task 5: improving aspect based sentiment analysis using neural network features», en *Proceedings of SemEval-2016*, San Diego, California, 2016, pp. 282-288.
[107] H. Wang, Y. Lu, y C. Zhai, *Latent aspect rating analysis on review text data: a rating regression approach*. ACM, 2010.
[108] H. Wang, Y. Lu, y C. Zhai, «Latent aspect rating analysis without aspect keyword supervision», en *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011, pp. 618-626.
[109] D.-H. Pham y A.-C. Le, «Determining aspect ratings and aspect weights from textual reviews by using neural network with paragraph vector model», en *International Conference on Computational Social Networks*, 2016, vol. 9795, pp. 309-320.
[110] D.-H. Pham y A.-C. Le, «Learning multiple layers of knowledge representation for aspect based sentiment analysis», *Data & Knowledge Engineering*, jun. 2017.
[111] Y. Ding, C. Yu, y J. Jiang, «A neural network model for semi-supervised review aspect identification», en *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2017, pp. 668-680.
[112] L. Xu, J. Liu, L. Wang, y C. Yin, «Aspect based sentiment analysis for online reviews», en *Advances in Computer Science and Ubiquitous Computing. CUTE 2017, CSA 2017. Lecture Notes in Electrical Engineering*, Taichung, Taiwan, 2017, vol. 474, pp. 475-480.
[113] D.-H. Pham y A.-C. Le, «Fine-tuning word embeddings for aspect-based sentiment analysis», en *International Conference on Text, Speech, and Dialogue*, 2017, pp. 500-508.

[114] X. Li y W. Lam, «Deep multi-task learning for aspect term extraction with memory interaction», en *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2876-2882.

[115] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, y P. Kuksa, «Natural language processing (almost) from scratch», *Journal of Machine Learning Research*, vol. 12, n.º Aug, pp. 2493-2537, 2011.

[116] Z. Toh y S. Jian, «NLANGP: supervised machine learning system for aspect category classification and opinion target extraction», en *Proceedings of the 9th International Workshop on Semantic Evaluation*, Denver, Colorado, USA, 2015, pp. 496-501.

[117] H. Xu, B. Liu, L. Shu, y S. Y. Philip, «Double embeddings and cnn-based sequence labeling for aspect extraction», en *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2018, vol. 2, pp. 592-598.

[118] M. Schmitt, S. Steinheber, K. Schreiber, y B. Roth, «Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks», en *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 1109-1114.

[119] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, y T. Mikolov, «Learning word vectors for 157 languages», en *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[120] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, y B. Qin, «Learning sentiment-specific word embedding for twitter sentiment classification», en *Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics*, 2014, vol. 1, pp. 555–1565.

[121] D. Tang, B. Qin, X. Feng, y T. Liu, «Effective lstms for target-dependent sentiment classification», en *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 3298-3307.

[122] D.-T. Vo y Y. Zhang, «Target-dependent twitter sentiment classification with rich automatic features», en *International Joint Conferences on Artificial Intelligence, IJCAI*, 2015, pp. 1347-1353.

[123] R. Ma, K. Wang, T. Qiu, A. K. Sangaiah, D. Lin, y H. B. Liaqat, «Feature-based compositing memory networks for aspect-based sentiment classification in social internet of things», *Future Generation Computer Systems*, 2017.

[124] D. Mimno, H. M. Wallach, E. Talley, M. Leenders, y A. McCallum, «Optimizing semantic coherence in topic models», en *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2011, pp. 262-272.

[125] H. Lakkaraju, R. Socher, y C. Manning, «Aspect specific sentiment analysis using hierarchical deep learning», en *NIPS Workshop on Deep Learning and Representation Learning*, 2014.

[126] B. Li, L. Zhou, S. Feng, y K.-F. Wong, «A unified graph model for sentence-based opinion retrieval», en *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 2010, pp. 1367-1375.

[127] A. Lazaridou, I. Titov, y C. Sporleder, «A bayesian model for joint unsupervised induction of sentiment, aspect and discourse representations», en *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2013, vol. 1, pp. 1630-1639.

[128] S. Gu, L. Zhang, Y. Hou, y Y. Song, «A Position-aware bidirectional attention network for aspect-level sentiment Analysis», en *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 774-784.

[129] C. Wu, F. Wu, S. Wu, Z. Yuan, y Y. Huang, «A hybrid unsupervised method for aspect term and opinion target extraction», *Knowledge-Based Systems*, vol. 148, pp. 66-73, 2018.

[130] W. Xue, W. Zhou, T. Li, y Q. Wang, «MTNA: a neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews», en *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 2017, vol. 2, pp. 151-156.

[131] W. Wang, S. J. Pan, D. Dahlmeier, y X. Xiao, «Coupled multi-layer attentions for co-extraction of aspect and opinion terms», en *Proceedings of the Thirty-First Conference on Artificial Intelligence*, 2017, pp. 3316-3322.

[132] J. Kirkpatrick *et al.*, «Overcoming catastrophic forgetting in neural networks», *Proceedings of the National Academy of Sciences*, vol. 114, n.º 13, pp. 3521-3526, 2017.

[133] D. Reforgiato Recupero, E. Cambria, y E. Di Rosa, «Semantic sentiment analysis challenge at eswc 2017», Cham, 2017, pp. 109-123.