

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/325370586>

A SURVEY ON ASPECT BASED SENTIMENT ANALYSIS

Conference Paper · March 2018

CITATIONS

0

READS

731

2 authors:



[Yisel Clavel](#)

University of Holguín

4 PUBLICATIONS 1 CITATION

[SEE PROFILE](#)



[Leticia Arco García](#)

Vrije Universiteit Brussel

63 PUBLICATIONS 233 CITATIONS

[SEE PROFILE](#)



Informática
XIII CONVENCIÓN Y FERIA INTERNACIONAL
13th INTERNATIONAL CONVENTION AND FAIR 2018

La Habana, Cuba. Del 19 al 23 de marzo
Havana, Cuba. March 19th to 23th

ESTUDIO DEL ANÁLISIS DE SENTIMIENTOS BASADO EN ASPECTOS

A SURVEY ON ASPECT BASED SENTIMENT ANALYSIS

Yisel Clavel Quintero¹, Leticia Arco García²

¹ Universidad de Holguín, Cuba, yclavelq@uho.edu.cu, Avenida XX Aniversario, Vía Guardalavaca, Piedra Blanca, Holguín, Cuba, CP: 80100.

² Universidad Central "Marta Abreu" de Las Villas, Cuba, leticiaa@uclv.edu.cu

RESUMEN: A partir del crecimiento y desarrollo de las tecnologías móviles, el uso de Internet y el comercio electrónico, las fuentes de opinión de la web como los portales de compras en línea, los foros de discusión, grupos, blogs, micro blogs y redes sociales se usan extensamente para compartir información, experiencia y opiniones. La minería de opinión o análisis de sentimientos es un área de la minería de textos que se nutre de otras áreas como la inteligencia artificial y el aprendizaje automático. La opinión es evaluada según su polaridad en positiva, negativa, neutral y conflictiva. En el análisis de sentimientos basado en aspectos se identifican los aspectos de una entidad que expresa sentimiento sobre un producto o servicio determinado, al nivel más profundo de análisis en una oración. El presente estudio presenta el estado del arte y los principales retos existentes en el análisis de sentimientos basado en aspectos de revisiones de productos y servicios.

Palabras claves: minería de opinión, análisis de sentimientos, análisis de sentimientos basado en aspectos, extracción de aspectos, aprendizaje profundo.

ABSTRACT: From the increase and development of the mobile technologies, the use of Internet and the ecommerce, web opinion sources like online shopping portals, discussion forums, peer-to-peer networks, groups, blogs, micro blogs and social networking applications are extensively used to share the information, experience and opinions. The opinion mining or sentiment analysis is the text mining area that feeds on other areas like artificial intelligence and machine learning. The opinion is evaluated to its positivity, negativity, neutrality and conflict. In the aspect based sentiment analysis, the aspects of an entity are identified. These aspects express sentiments on a product or service, a more fine-grained analysis in a sentence. The present survey shows the state-of-the-art and the main existing challenges in the aspect based sentiment analysis of product and service reviews.

Keywords: opinion mining, sentiment analysis, aspect based sentiment analysis, aspect extraction, deep learning.

1. INTRODUCCIÓN

A partir del desarrollo de las tecnologías móviles, el uso de Internet y el comercio electrónico; se han hecho populares las fuentes de opinión en la web como los portales de compras en línea, los foros de discusión, blogs, redes sociales y sitios de recomendaciones. Estas fuentes se usan para compartir información, experiencia y opiniones.

Las opiniones en el mundo empresarial tienen gran importancia debido a que para una empresa resulta vital conocer lo que piensan los usuarios sobre las virtudes y debilidades de sus productos o servicios. Por eso, es indispensable el desarrollo de técnicas que permitan analizar las opiniones y arribar a conclusiones que le permitan a la empresa tomar las decisiones adecuadas: bajar los precios, cambiar componentes, rediseñar el producto, etc., [1].

Las opiniones son los estados subjetivos que reflejan los sentimientos y la percepción de una persona sobre un suceso o un objeto. Una opinión es una evaluación sobre una entidad o un aspecto de la entidad expresada por un usuario [2].

La minería de opinión o análisis de sentimientos es el estudio computacional de opiniones, sentimientos y emociones expresados en textos. Es un área de la minería de textos que abarca técnicas del procesamiento del lenguaje natural (*Natural Language Processing*; NLP) y la lingüística computacional, que tienen como objetivo la extracción de información subjetiva a partir de contenidos generados por los usuarios [3]. Consiste en la clasificación de palabras, textos o documentos de acuerdo a las opiniones, sentimientos, emociones y subjetividades expresadas [4].

Acerca del análisis de sentimiento basado en aspectos (*Aspect Based Sentiment Analysis*; ABSA) se han realizado varias investigaciones en la última década que proponen varios enfoques y técnicas [5, 6]. Por esta razón, surge la necesidad de sintetizar las principales características y técnicas empleadas.

En [6] se mencionan algunas herramientas usadas en cada tarea de ABSA y se realiza una evaluación de los resultados de diferentes artículos en conjuntos de datos de restaurantes y *laptops*, pero no se caracterizan ni comparan los resultados de dichas herramientas. En [7] se caracterizan los distintos enfoques y tipos de aprendizaje aplicados a ABSA, así como las métricas usadas para la evaluación, pero no se realiza un estudio de las herramientas más usadas. En [6] se destaca el aprendizaje profundo (*deep learning*) como una técnica que puede ser aplicada con éxito en ABSA, pero no se muestran estudios de su aplicación. En [5] y [8] se describen los distintos niveles de análisis en la minería de opinión, se mencionan los enfoques y las técnicas más usadas en cada nivel y se muestran los principales recursos empleados. En [5] se presentan algunos trabajos de aprendizaje profundo para el desarrollo del NLP y la minería de opinión. Los estudios analizados no mencionan técnicas de análisis de sentimiento de múltiples aspectos.

El artículo tiene como objetivo caracterizar el análisis de sentimientos basado en aspectos, a partir de un análisis comparativo de las principales técnicas empleadas en cada etapa, destacando nuevas técnicas como el aprendizaje profundo y la clasificación multi-aspectos, así como identificar los principales retos.

2. PRINCIPALES DEFINICIONES DEL ANÁLISIS DE SENTIMIENTOS

Las principales tareas del análisis de sentimiento comprenden la extracción y categorización de entidades, aspectos, titulares de opinión, la extracción

y estandarización del tiempo de la opinión, y la clasificación de la polaridad [2]. En la clasificación de la polaridad, la opinión es evaluada según su polaridad en: positiva, negativa, neutral y conflictiva. Algunos enfoques utilizan la multipolaridad, que clasifica la polaridad en: fuertemente positiva, débilmente positiva, débilmente negativa, y fuertemente negativa [6].

La clasificación del sentimiento o clasificación de la polaridad es una tarea del aprendizaje automático que puede ser supervisada, no supervisada o semi-supervisada. Los clasificadores que siguen un enfoque supervisado requieren de colecciones textuales previamente etiquetadas y se caracterizan por ofrecer un buen rendimiento en base al dominio en el que son entrenados. Este tipo de aprendizaje ha obtenido los mejores resultados para la clasificación de la polaridad o categorías de aspectos. El aprendizaje semi-supervisado [9] consiste en usar un conjunto de ejemplos sin etiqueta para auxiliar el aprendizaje supervisado para la clasificación basado en una pequeña cantidad de ejemplos etiquetados [10]. Los clasificadores no supervisados no necesitan colecciones etiquetadas para su funcionamiento, no cuentan con categorías o atributos objetivos previamente definidos [11] y pueden ser empleados en diferentes dominios con buen rendimiento, por lo que han sido utilizados con éxito en la extracción de aspectos [5].

Por otra parte, existen varios niveles de análisis en la minería de opinión: documento, oración, frase y aspecto. En el nivel de documento se determina si toda la opinión que se plasma en un documento expresa un sentimiento positivo, negativo, neutral o conflictivo. Para ello, se clasifica cada documento según su polaridad. Este nivel supone que cada documento expresa opiniones sobre un solo objeto o entidad y contiene la opinión de un solo titular de opinión. Por esta razón, este nivel no es aplicable a blogs o foros u otras fuentes de datos donde existan múltiples opiniones de diferentes objetos [12, 13].

En el nivel de oración se determina el sentimiento que se expresa en una oración (positivo, negativo, neutral o conflictivo). Usualmente, este es un paso intermedio en las distintas tareas de análisis de sentimiento. La polaridad de cada oración generalmente se puede calcular con los mismos métodos del nivel de documento [2]. Generalmente se asume que cada oración expresa una sola opinión, aunque algunas veces esto no es cierto. Esta suposición es apropiada para oraciones simples, pero oraciones compuestas o complejas pueden expresar más de una opinión [2].

El nivel de frase en la clasificación de sentimiento es un enfoque mucho más preciso para la minería de opinión, donde se clasifica el sentimiento de una expresión. Las frases que contienen palabras de opinión deben ser detectadas y luego clasificadas según su polaridad. En ocasiones la polaridad

depende del contexto, por lo que es necesario analizar el rol de las palabras de negación y frases que comienzan con *but* en inglés (en español: excepto, pero, no obstante, sin embargo, etc.) [14]. Los enfoques existentes en este nivel están usualmente basados en lexicones, corpus etiquetados según la polaridad [15] o en aspectos obtenidos a partir de léxicos generados manualmente o automáticamente [16]. Otros, emplean el análisis gramatical de dependencia de frases y construyen un árbol de dependencia de frases [17].

Los aspectos o características son definidos como atributos o componentes de objetos. El análisis de éstos mejora la clasificación de la polaridad del documento u objeto respecto a los niveles anteriores; ya que realiza un análisis con un nivel de granularidad más fino, siendo el nivel más profundo de análisis en un documento. Este nivel tiene como objetivo descubrir sentimientos sobre entidades y/o sus aspectos. Inicialmente este tipo de análisis fue llamado análisis a nivel de rasgos (minería de opinión basada en rasgos y sumarización). No solo permite determinar la subjetividad u opinión de un objeto sino qué le gusta o no del objeto al cliente. En lugar de buscar sobre constructores del lenguaje (documentos, párrafos, oraciones, o frases), busca directamente sobre la opinión en sí. Se basa en que una opinión consiste de un sentimiento y una etiqueta (de opinión) [1]. Los aspectos deben ser extraídos y clasificados según su polaridad. Luego, estos deben ser agrupados para crear un resumen de opinión basado en aspectos de múltiples documentos [12, 13].

3. ANÁLISIS DE SENTIMIENTOS BASADO EN ASPECTOS

El principal objetivo del Análisis de Sentimientos Basado en Aspectos es identificar los aspectos de una entidad y el sentimiento expresado por cada aspecto sobre un producto o servicio determinado. Primeramente se descubre el objeto sobre el cual son expresadas las opiniones en una oración y luego se determina la polaridad de las opiniones. Este tipo de análisis es bastante usado en la minería y resumen automático de opiniones. En la última década se han desarrollado *software* que realizan ABSA y han sido evaluados con artículos de revisiones de películas, viajes, cámaras digitales, computadoras, restaurantes, entre otros productos y servicios [6].

Un objeto puede ser un producto, servicio, individuo, organización, evento, tópico, etc. Por ejemplo: en la oración "el tiempo de vida de la batería de este teléfono es muy corto"; el comentario es sobre "el tiempo de vida de la batería", donde "batería" es el aspecto, "teléfono" es el objeto o entidad y la opinión es negativa. Muchas aplicaciones del mundo real necesitan este nivel de análisis para hacer mejoras a los productos y servicios, a partir de identificar qué

componentes y/o características de los productos o servicios les gustan o no a los clientes [3].

En [6] se divide el proceso en dos fases: (1) la extracción de términos aspectos y agrupar los aspectos en categorías; y (2) la identificación de la polaridad de los términos aspectos y la polaridad de las categorías de aspectos de cada oración. Estas fases de actividades están divididas en cuatro subtareas principales [6, 18]: Extracción de Términos Aspectos (*Aspect Term Extraction*; ATE); Polaridad de Términos Aspectos (*Aspect Term Polarity*; ATP); Detección de Categorías de Aspectos (*Aspect Category Detection*; ACD); y Polaridad de Categorías de Aspectos (*Aspect Category Polarity*; ACP) [6, 19]. En [20] y [21] se definen tres subtareas: extracción de la categoría de aspecto, identificación del objetivo de la opinión y cálculo de la polaridad. En [5] se definen dos subtareas para ABSA: la extracción de aspectos y la clasificación del sentimiento. En [22] también se proponen dos subtareas con el mismo enfoque, solo que las nombran de manera diferente: identificación del aspecto y predicción de la polaridad del aspecto. En este artículo se abordan las cuatro subtareas empleadas en [18].

3.1 Extracción de Términos Aspectos

El objetivo de esta tarea es extraer el término aspecto, por tanto, se considera una tarea de extracción de información. Un aspecto puede ser expresado por un sustantivo, adjetivo, adverbio y verbo. La mayoría de los términos aspectos (del 60% al 70%) son sustantivos. Los aspectos pueden ser además entidades multi-palabras (*n-gram*) como "batería de respaldo" ("*battery backup*") [6].

Algunos de los enfoques de esta tarea están basados en la frecuencia; es decir, se basan en la presencia o ausencia de términos; algunos aceptan los aspectos más frecuentes, otros eliminan los sustantivos y frases sustantivas con mayor ocurrencia que pueden no ser aspectos [12]. Otros enfoques usan aprendizaje automático o modelado de tópicos [8, 23]; otros se basan en reglas para extraer los sustantivos, adjetivos, adverbios y verbos [6].

En esta subtarea se emplean técnicas de procesamiento del lenguaje natural y extracción de información [6] para el preprocesamiento del texto, como: tokenización [14, 24], normalización de mayúsculas y minúsculas [14], eliminación de palabras vacías (*stop-words*) [6, 11, 14], *stemming* [11, 14, 25], etiquetado de las partes de la oración (*part-of-speech*; POS) [5], la extracción de entidades nombradas [12, 26], la extracción de fragmentos significativos del texto (*chunking*) [25], la segmentación de las oraciones [26], la construcción de *n-grams* a partir de palabras consecutivas [27], la desambiguación del sentido de las palabras [25, 28] y la resolución de co-referencias [26, 28]. Algunas

técnicas¹ para realizar esta tarea se muestran en la Tabla 1, así como el tipo de aprendizaje de cada una.

Tabla 1. Técnicas para la extracción de términos aspectos.

Técnica	Tipo de aprendizaje
Latent Dirichlet Allocation (LDA) [29]	no supervisado
Pointwise Mutual Information (PMI) [30]	no supervisado
Double Propagation Algorithm [31]	no supervisado
Association Mining [32, 33]	no supervisado
Deep learning [34]	no supervisado
Recurrent Neural Networks (RNN) [5]	no supervisado
Convolutional Neural Networks (CNN) [35]	no supervisado
Latent Semantic Analysis (LSA) [14]	semi- supervisado
Maximum-Entropy (ME) [36]	supervisado
Support Vector Machines (SVM) [37]	supervisado
Conditional Random Fields (CRF) [38]	supervisado
Rule-based methods [39]	supervisado

Uno de los procedimientos no supervisados de modelado de tópicos más conocidos es el denominado *Latent Dirichlet Allocation* (LDA) [29], que permite analizar la semántica de la distribución de temas en documentos. El modelo LDA define un proceso generativo probabilístico de Dirichlet para la distribución de documentos-temas. En cada documento, un aspecto latente se elige de acuerdo con una distribución multinomial, controlada por una prioridad de Dirichlet; entonces, dado un aspecto, una palabra se extrae de acuerdo con otra distribución multinomial, controlada por otra prioridad de Dirichlet [40]. Aunque existen implementaciones fáciles de utilizar, tiene el inconveniente de que es preciso elegir de antemano el número de tópicos deseados, lo cual no siempre es sencillo [41].

En [40] se presenta un nuevo marco de trabajo denominado Sentic LDA, enfoque estadístico que integra la computación de sentido común en el cálculo de las distribuciones de palabras en el algoritmo LDA, lo que permite el cambio de sintaxis a semántica en el análisis de sentimiento basado en aspectos. Sentic LDA extrae los aspectos aprovechando la semántica asociada con palabras y agrupa con éxito los términos aspectos según su categoría de aspectos.

En [9] se propusieron dos modelos estadísticos semi-supervisados, *Seeded Aspect and Sentiment* (SAS) y *Maximum Entropy - Seeded Aspect and Sentiment* (ME-SAS), basados en LDA, para extraer y categorizar aspectos automáticamente a partir de palabras semillas que reflejan las categorías de interés del usuario y sus necesidades para descubrir aspectos específicos. ME-SAS mejora SAS usando Maximización de la Entropía (*Maximum-Entropy*; ME). Por otra parte, en [36] se emplea ME para entrenar una variable de conmutación basada en las etiquetas POS de las palabras y usarla para separar aspectos y palabras de sentimiento.

Varios enfoques de modelado de tópicos basados en LDA han sido propuestos para el análisis de sentimientos multi-aspectos [36, 42, 43], que persigue tener en cuenta diversos aspectos potencialmente relacionados y a menudo discutidos dentro de un mismo artículo de opinión [23]. En [23] se muestra que los modelos semi-supervisados se desempeñan bastante bien en tareas de etiquetado de oraciones con múltiples aspectos, y también pueden ayudar a la predicción de la evaluación de múltiples aspectos solo con supervisión indirecta para predecir clasificaciones de estrellas específicas asociadas a aspectos implícitos para cada revisión; con rendimiento comparable al clasificador supervisado Máquina de Soporte Vectorial (*Support Vector Machine*; SVM). Además, los autores afirman que incorporar características derivadas de modelos no supervisados proporciona un sustancial aumento del rendimiento en modelos débiles como *Perceptron Rank*; y en menor medida en modelos más fuertes como *Support Vector Regression* [23].

En [30] se introduce el sistema OPINE que extrae aspectos explícitos basado en *Pointwise Mutual Information* (PMI), usada para estimar la orientación semántica de las frases de opinión, asignando una puntuación entre hechos y frases discriminadoras o modificadoras [37, 44]. PMI, en otras palabras, mide la asociación de palabras basada en la teoría de la información [45] y determina los aspectos probables asignando una puntuación entre las frases sustantivas y los discriminadores de meronimia asociados con la clase del producto (por ejemplo: las frases “*of pone*”, “*phone has*”, “*phone comes with*”, para la clase *Phone*) [5].

En [31] se presentó un algoritmo de propagación doble para descubrir pares de palabra de opinión y el objetivo de la opinión. Las palabras de opinión se usan para encontrar más objetivos de opinión y, a su vez, los objetivos de opinión se utilizan para encontrar más palabras de opinión. Sin embargo, este algoritmo no funciona bien sobre conjuntos de datos muy grandes o muy pequeños [5, 46].

En [46] se propusieron dos mejoras del algoritmo de propagación doble que logra encontrar más aspectos basándose en patrones de negación y de la forma parte-todo. Se usan los patrones parte-todo (es decir, discriminadores de meronimia) para indicar si un aspecto es parte de una entidad o no. Un algoritmo de ordenamiento jerárquico de aspectos se aplica para encontrar los aspectos más probables determinados por la relevancia y frecuencia del aspecto. El algoritmo de propagación doble realiza la extracción de aspectos y su ordenamiento, asumiendo que los aspectos son sustantivos y frases sustantivas y las palabras de opinión son adjetivos.

La Minería de Asociación [33] es adoptada para detectar elementos frecuentes en las revisiones de productos, y las técnicas de poda son usadas para

¹ Mantuvimos sus nombres originales en inglés para su fácil localización en la literatura científica.

reducir aspectos no deseados. En [33] se asume que los aspectos son sustantivos o frases sustantivas; por tanto, antes de detectar los aspectos se aplica algún etiquetador POS.

En [47] se presenta un enfoque de minería de asociación para extraer aspectos explícitos e implícitos. Las palabras de opinión se agrupan según su polaridad en positivas o negativas. Los autores afirman que una opinión implícita y la opinión explícita vecina tienden a abordar el mismo aspecto y la misma polaridad. A partir de esto, construyen automáticamente un corpus anotado de opiniones implícitas con clases de aspectos y polaridad, luego el corpus se utiliza para entrenar clasificadores de aspectos y polaridad que podrán reconocer los aspectos y la polaridad de opiniones implícitas [37].

Una arquitectura profunda es formada por la composición de varios niveles de representación, donde la cantidad de niveles es un parámetro libre, el cual puede seleccionarse dependiendo de la tarea a ejecutar [48]. El aprendizaje profundo es un tipo de enfoque con múltiples niveles de representación del aprendizaje, el cual se ha empleado en aplicaciones de visión por computadoras, reconocimiento de lenguaje y NLP [5]. El reuso, que constituye el poder de las representaciones distribuidas, es la principal ventaja teórica del aprendizaje profundo, así como construir niveles múltiples de representación o aprender una jerarquía de características. Otra importante motivación para aplicar el aprendizaje profundo en la extracción de aspectos es que algunas de sus técnicas no requieren ejemplos previamente etiquetados, por lo que están relacionadas con los principios del aprendizaje no supervisado. Además, puede explotarse de forma semi-supervisada (donde solo unos pocos ejemplos son etiquetados) [48]. En [34] se diseñó un marco de trabajo para el aprendizaje profundo que permite extraer los aspectos y su sentimiento asociado [49, 50].

En [51] se propuso una clase general de modelos discriminativos basados en Redes Neuronales Recurrentes (*Recurrent Neural Networks*; RNN) y palabras embebidas que pueden ser exitosamente aplicados en la minería de opinión basada en aspectos. Una RNN es una red neuronal que puede manejar secuencias de entrada de longitud variable [5]. En cierto sentido, las RNN son las más profundas y poderosas de todas las redes neuronales, y pueden en principio crear y procesar secuencias de memorias arbitrarias de patrones de entrada. A diferencia de los métodos tradicionales para síntesis automática de programas secuenciales, las RNN pueden aprender programas que mezclan procesamiento de información secuencial y paralela de manera natural y eficiente, explotando el paralelismo masivo [52].

Por otro lado, las Redes Neuronales Convolucionales (*Convolutional Neural Networks*; CNN) [35], son empleadas para el reconocimiento de la polaridad y

los aspectos implícitos, con buenos resultados. Utilizan capas con múltiples filtros que son aplicados para obtener múltiples aspectos.

En [49] se ha introducido un enfoque basado en el aprendizaje profundo para la extracción de aspectos, con una significativa mejora en rendimiento sobre los demás enfoques del estado del arte. Se asume que las características de un término de aspecto dependen de sus palabras circundantes. Se propone una arquitectura de CNN profunda específica que comprende siete capas: la capa de entrada, que consiste en características de palabras incluidas por cada palabra en la oración; dos capas de convolución, cada una seguida por una capa de acumulación máxima; una capa completamente conectada; y, finalmente, la capa de salida, que contiene una neurona por cada palabra. En [34] se diseñó un modelo de aprendizaje profundo para analizar los sentimientos basados en aspectos que muestra un rendimiento competitivo o mejor en comparación con los mejores resultados de SemEval 2015 [20] en todas las sub tareas.

El Análisis de Latencia Semántica (*Latent Semantic Analysis*; LSA) es una técnica estadística para descubrir automáticamente una estructura semántica, es decir, encontrar asociaciones semánticas entre los términos en una colección de documentos [14] y tiene una alta complejidad computacional.

En [53] se centran exclusivamente en identificar qué expresión de opinión se relaciona con qué aspecto en una oración de opinión sobre un producto. Para ello, entrenan un clasificador SVM para encontrar la expresión de opinión relacionada y el aspecto objetivo. El objetivo de esta propuesta es aprender un modelo que ordena jerárquicamente los aspectos que ocurren en la misma oración como una expresión de opinión tal que los que están mejor posicionados son los más probables a ser el objetivo. Los vectores de características se forman en base a la relación sintáctica y semántica entre la expresión de opinión y el aspecto candidato [8].

Los Campos Aleatorios Condicionales (*Conditional Random Fields*; CRF) [38] son un tipo de modelo gráfico probabilístico discriminativo no dirigido utilizado para codificar relaciones conocidas entre las observaciones y construir interpretaciones coherentes. Se utiliza principalmente para el etiquetado secuencial de las entidades. Un CRF de cadena lineal impone restricciones a las relaciones entre entidades, tales que cada entidad solo está vinculada a su entidad inmediatamente anterior y posterior, formando así una cadena lineal [54]. En [54] se emplea CRF con un árbol de análisis de dependencias, un etiquetador POS, la similitud coseno, los términos *unigrams* y el recurso SentiWordNet² para categorizar una palabra como

² <http://sentiwordnet.isti.cnr.it/>

palabra de opinión; con un desempeño peor que las demás técnicas analizadas en [20]. En [55] y en [56] se representa el texto como un vector de características y se trata como una tarea de etiquetado de secuencia, donde cada *token* en una oración candidata está etiquetado como *Beginning*, *Inside or Outside* (BIO); y se emplea CRF para estimar la probabilidad condicional de una etiqueta de secuencia a partir de sus características. Para la extracción de aspectos, CRF [57, 58] obtuvo los mejores resultados en SemEval 2014 [18], en dominios de restaurantes ($F1=84.01\%$) y laptops ($F1=74.55\%$), con etiquetado POS y utilizando el árbol de dependencias.

En [39] se propone un método de extracción de entidades y aspectos basado en reglas. Múltiples reglas lingüísticas, las cuales son representadas como expresiones regulares de etiquetas POS, son diseñadas para extraer los sustantivos adjetivales, los sustantivos adverbiales y las frases sustantivas extendidas. Los métodos basados en reglas son efectivos en la detección de entidades y aspectos explícitos [39]. Las oraciones comparativas pueden ser empleadas para la detección de aspectos implícitos [5, 59].

3.2 Detección de la Polaridad de Términos Aspectos

En esta tarea, dado un conjunto de términos aspectos, el objetivo es determinar la polaridad de cada aspecto en: positivo, negativo, neutral o conflictiva. Para esto se emplean técnicas de clasificación. Son muy usadas la polaridad de adjetivos vecinos, las etiquetas POS vecinas y las relaciones de dependencia entre los términos [6]. La Tabla 2 muestra algunas de las técnicas empleadas para la detección de la polaridad de términos aspectos.

Tabla 2. Técnicas para la detección de la polaridad de términos aspectos.

Técnica	Tipo de aprendizaje
K mean [60]	no supervisado
Double Propagation Algorithm [31]	no supervisado
Expectation Maximization (EM) [25]	semi- supervisado
Naive Bayes [61]	supervisado
Support Vector Machines (SVM) [37]	supervisado
Hidden Markov Model (HMM) [62]	supervisado
Random Forest [63]	supervisado
Neural Network [64]	supervisado y no supervisado

La técnica de agrupamiento K-mean empleada en [60], divide n aspectos en k grupos calculando la distancia mínima entre el valor de sentimiento de los aspectos y el valor medio de los grupos. Los aspectos se agrupan en positivos fuertes, positivos débiles, negativos débiles y negativos fuertes.

El clasificador *Naive Bayes* es un modelo probabilístico muy conocido y con una implementación muy simple. *Naive Bayes* trata de estimar la probabilidad de que un texto pertenezca a una categoría [65] y típicamente trabaja con representación de bolsas de palabras [25].

Por otro lado, *Expectation Maximization* (EM) es una técnica general que trabaja sobre una combinación de datos supervisados y no supervisados. Por ejemplo, el método de agrupamiento EM se ha combinado de manera exitosa con el clasificador *Naive Bayes*. El algoritmo EM computa iterativamente las expectativas (en el paso E) dado un modelo actual, usándolas como datos de entrenamiento para maximizar un modelo de estimación (en el paso M). La combinación del paso E con el M garantiza reducir errores. Esta técnica puede trabajar con ejemplos no etiquetados [25].

En [66] se usan métodos de aprendizaje supervisado como *Naive Bayes* y SVM para clasificar revisiones de películas [44]. Las SVM con núcleo lineal son robustas en el reconocimiento de la polaridad implícita [37]. El mejor resultado para la detección de la polaridad de aspectos en SemEval 2014 [18] fue usando SVM ($accuracy=80.95\%$ en revisiones de restaurantes) [67], con características basadas en n -grams y utilizando los árboles sintácticos, el reconocimiento de entidades nombradas, un etiquetador POS [54] y los lexicones de sentimientos MPQA³, SentiWordNet y Bing Liu's Opinion Lexicon⁴. En [62] se introdujo el marco de trabajo OpinionMiner basado en los Modelos Ocultos de Markov (*Hidden Markov Model*; HMM). OpinionMiner combina múltiples características lingüísticas significativas (por ejemplo: etiquetas POS, patrones de información interna de frases y claves contextuales circundantes) en un proceso de aprendizaje automático. Para etiquetar los datos de entrenamiento, los autores diseñaron un enfoque que podría extraer datos etiquetados de alta confianza a través del autoaprendizaje [8].

En [63] se desarrolla un método de clasificación *Random Forest* independiente del dominio, pero que obtiene resultados discretos comparados con los obtenidos por las técnicas analizadas en [18]. En [68] el sentimiento del aspecto es computado usando el sentimiento de cada n -gram y la distancia entre el n -gram y el aspecto.

En [64] se propone un sistema basado en una arquitectura de redes neuronales en dos pasos para la extracción de aspectos y el sentimiento asociado. Se emplea RNN para realizar la extracción de aspectos como una tarea de etiquetado de secuencia. Se convierte a minúsculas el texto y se mantienen las palabras gramaticales y los signos de puntuación. Además, se emplean características derivadas de dos

³ http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/

⁴ <http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar>

recursos semánticos basados en grafos: WordNet⁵ y SenticNet [69] (proporciona puntuaciones de valor real para cinco *sentic*s: *pleasantness*, *attention*, *sensitivity*, *aptitude*, *polarity*). Para predecir la etiqueta de polaridad para un aspecto específico en una oración, se marcan los aspectos ya extraídos y se etiqueta cada palabra de la oración con su distancia relativa al aspecto. Para el etiquetado, la secuencia de palabras se mapea a una secuencia de vectores de palabras, vectores *sentic* y vectores POS. Luego, se concatena cada vector de palabra con su vector *sentic* correspondiente, su vector de etiqueta POS y el vector de distancia. Se considera la etiqueta con la mayor probabilidad estimada de ser la etiqueta de polaridad predicha (positivo o negativo) para el término de aspecto dado. Esta propuesta puede detectar múltiples opiniones en una oración y obtuvo los mejores resultados en 2016 ESWC fine-grained sentiment analysis challenge (accuracy = 87.4%) [64].

3.3 Detección de Categorías de Aspectos

Esta tarea es donde se identifican las categorías de aspectos discutidas en cada oración. Las categorías de aspectos son típicamente más generales que los términos aspectos extraídos, y no necesariamente ocurren como términos en la oración [6]. Los aspectos extraídos en la tarea 1 deben ser agrupados en categorías de aspectos, que pueden ser basadas en aspectos sinónimos; por ejemplo, "calidad de la llamada" y "calidad de la voz" se refieren al mismo aspecto de la entidad teléfono. Las categorías de aspectos pueden no ocurrir como términos en la oración; por ejemplo, en la oración "deliciosos pero caro" las categorías "comida" y "precio" no aparecen como términos [18].

Esta tarea puede ser considerada como un problema de clasificación multi-etiqueta, tarea crítica del análisis de opinión, ya que muchos sinónimos son dependientes del dominio. Por ejemplo, "movie" (película) y "picture" (fotografía) son sinónimos en revisiones de películas, pero no en revisiones de cámara donde "picture" es sinónimo de "photo" (foto) y "movie" de "video" (vídeo) [2].

En [20] se propone como categoría de aspecto una combinación de una entidad *E* y su atributo *A*. Por lo que el objetivo de esta tarea es identificar cada par (*E*, *A*) sobre el cual es expresada una opinión en el texto dado. Para la detección de la categoría se emplean algoritmos de clasificación como los que se muestran en la Tabla 3.

En [68] un algoritmo basado en co-ocurrencias de términos se usa en la detección de categorías. Éste obtiene una matriz de co-ocurrencias que captura la frecuencia de las co-ocurrencias entre palabras en la oración y la categoría de aspecto anotada; brindando

una correspondencia de palabras a categorías de aspectos.

Tabla 3. Técnicas para la detección de categorías

Técnica	Tipo de aprendizaje
Algoritmo basado en co-ocurrencia [68]	-
Binary Maximum Entropy (ME) [56]	supervisado
Random Forest [70]	supervisado
SVM [71]	supervisado

En [56] se emplea un conjunto de clasificadores binarios ME, uno por cada clase, para la detección de la categoría de aspectos. En [70] se identifica cada par (*E*, *A*) y se aplican clasificadores *Random Forest* para cada par (*E*, *A*) de categorías de aspectos en los datos de entrenamiento. Para esta tarea, SVM [71] obtuvo los mejores resultados en SemEval 2014 ($F1=88.57\%$ en revisiones de restaurantes) [18].

3.4 Detección de la Polaridad de Categorías de Aspectos

Esta tarea toma las categorías detectadas en la tarea anterior y determina la polaridad de cada categoría de aspecto discutida en la oración [6]. Algunas de las técnicas que se emplean para esta tarea se muestran en la Tabla 4.

Tabla 4. Técnicas para la detección de la polaridad de las categorías

Técnica	Tipo de aprendizaje
Maximum Entropy (ME) [56]	supervisado
Stochastic Gradient Descent	supervisado
Adaboost	supervisado
SVM [71]	supervisado

Un clasificador ME es empleado para la detección de la polaridad en [56], el cual trabaja muy bien en muchas tareas de NLP. Otros enfoques en esta tarea calculan la distancia entre un término *n-gram* y el aspecto correspondiente [55, 68]. Para esta tarea, SVM [71] obtuvo los mejores resultados en SemEval 2014 (accuracy=82.92% en revisiones de restaurantes) [18].

4. RETOS

Como se puede apreciar en la bibliografía referenciada, el análisis de sentimiento basado en aspectos ha sido ampliamente abordado por la comunidad científica que desarrolla la minería de opinión. A partir del análisis realizado, se han identificado los siguientes retos en los que se debe seguir investigando para mejorar el desempeño de las técnicas empleadas en esta tarea:

- Mejorar el rendimiento del análisis de oraciones complejas y con múltiples opiniones.
- Identificar la polaridad implícita y los aspectos implícitos.

⁵ <http://wordnet.princeton.edu>

- Extraer los aspectos multi-palabras.
- Descubrir otras aplicaciones de las técnicas para el análisis de sentimiento multi-aspecto.
- Descubrir nuevas aplicaciones del aprendizaje profundo para la minería de opinión basada en aspectos.
- Mejorar el desempeño en la tarea de identificar relaciones semánticas y de dependencia entre los aspectos y el sentimiento asociado.
- Descubrir nuevas técnicas o combinaciones de éstas que favorezcan el análisis de sentimiento basado en aspectos en diferentes dominios.

5. CONCLUSIONES

El presente artículo caracteriza el análisis de sentimientos basado en aspectos a partir de sus principales definiciones, tareas, y un análisis comparativo de enfoques y técnicas; y muestra los principales retos en los que se puede seguir trabajando para mejorar el rendimiento del análisis de sentimiento.

El análisis de sentimiento basado en aspectos es el más profundo análisis realizado en un texto para identificar la polaridad o sentimiento del mismo, por lo que la aplicación de técnicas de este tipo favorece la precisión de los algoritmos de minería de opinión.

El análisis de sentimiento basado en aspectos cuenta con cuatro tareas principales: Extracción de Términos Aspectos, Polaridad de Términos Aspectos, Detección de Categorías de y Polaridad de Categorías de Aspectos; las cuales emplean varios enfoques y diversas técnicas de procesamiento de lenguaje natural y aprendizaje automático.

El aprendizaje no supervisado es el más empleado en la tarea de Extracción de Términos Aspectos, mientras que en las demás tareas predomina el uso de enfoques supervisados; destacándose CRF, RNN y SVM como las técnicas con mejores resultados obtenidos.

6. REFERENCIAS BIBLIOGRÁFICAS

1. Arco García, L., *Análisis de Sentimientos: Características Principales y Retos*. RECPAT 2015, 2015.
2. Liu, B., *Sentiment Analysis and Opinion Mining. Synthesis lectures on human language technologies*. 2012: Morgan & Claypool Publishers.
3. Liu, B., *Sentiment analysis and subjectivity*, in *Handbook of natural language processing*. 2010.
4. Amores, M., L. Arco, and M. Artiles, *PosNeg opinion: Una herramienta para gestionar comentarios de la Web*. Revista Cubana de Ciencias Informáticas, 2015. **9**(1).
5. Sun, S., C. Luo, and J. Chen, *A Review of Natural Language Processing Techniques for Opinion Mining Systems*. 2016.
6. Kumar Laskari, N. and S. Kumar Sanampudi, *Aspect Based Sentiment Analysis Survey*. IOSR Journal of Computer Engineering (IOSR-JCE), 2016. **18**(2).
7. C, T., S. R., and R. P., *A Comprehensive Survey on Aspect Based Sentiment Analysis*. International Journal of Advanced Research in Computer Science and Software Engineering, 2016. **6**(4).
8. Mohammad Sadegh, H., R. Ibrahim, and Z.A. Othman, *Opinion Mining and Sentiment Analysis: A Survey*. International Journal of Computers & Technology, 2012. **2**.
9. Mukherjee, A. and B. Liu, *Aspect Extraction through Semi-Supervised Modeling*, in *50th Annual Meeting of the Association for Computational Linguistics*. 2012, Association for Computational Linguistics: Jeju, Republic of Korea.
10. Poria, S., et al., *Fuzzy Clustering for Semi-Supervised Learning-Case study: Construction of an Emotion Lexicon*. 2013.
11. Huete Guadix, J.F. and J.M.F. Luna, *Indexación de documentos*, in *Recuperación de Información*. 2008, Universidad de Granada.
12. Mishra, N. and C.K. Jha, *An Insight into Task of Opinion Mining*. Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 2014.
13. Buche, A., M.B. Chandak, and A. Zadgaonkar, *Opinion Mining and Analysis: A Survey*. International Journal on Natural Language Computing (IJNLC), 2013. **2**(3).
14. Angulakshmi, G. and R. Manicka Chezian, *An Analysis on Opinion Mining: Techniques and Tools*. International Journal of Advanced Research in Computer and Communication Engineering, 2014. **3**(7).
15. Nam, S.-H., et al., *Partially Supervised Phrase-Level Sentiment Classification*. Springer-Verlag Berlin Heidelberg, 2009.
16. Townsend, R., et al., *WarwickDCS: From Phrase-Based to Target-Specific Sentiment Recognition*, in *9th International Workshop on Semantic Evaluation (SemEval 2015)*. 2015, Association for Computational Linguistics: Denver, Colorado.
17. Wu, Y., et al., *Phrase Dependency Parsing for Opinion Mining*, in *Conference on Empirical Methods in Natural Language Processing*. 2009, ACL and AFNLP: Singapore.
18. Pontiki, M., et al. *SemEval-2014 Task 4: Aspect Based Sentiment Analysis*. in *The 8th International Workshop on Semantic Evaluation (SemEval 2014)*. 2014. Dublin, Ireland.
19. Brun, C., D.N. Popa, and C. Roux, *XRCE Hybrid Classification for Aspect-based Sentiment Analysis*, in *8th International Workshop on Semantic Evaluation (SemEval 2014)*. 2014. Dublin, Ireland.
20. Pontiki, M., et al. *SemEval-2015 Task 12: Aspect Based Sentiment Analysis*. in *9th International*

- Workshop on Semantic Evaluation (SemEval 2015)*. 2015. Denver, Colorado: Association for Computational Linguistics.
21. Pontiki, M., et al. *SemEval-2016 Task 5 Aspect Based Sentiment Analysis*. in *SemEval-2016*. 2016. San Diego, California: Association for Computational Linguistics.
22. Luo, W., et al., *QPLSA Utilizing quad-tuples for aspect identification and rating*. Information Processing and Management, 2015. **51**.
23. Lu, B., et al., *Multi-aspect Sentiment Analysis with Topic Models*. 2011.
24. Arco García, L., *Agrupamiento basado en el concepto de intermediación diferencial y la aplicación de la teoría de los conjuntos aproximados para valorar resultados de agrupamientos*, in *Departamento de Ciencia de la Computación*. 2008, Universidad Central "Marta Abreu" de Las Villas: Santa Clara.
25. Alias-i. *LingPipe API Tutorials*. 2011 [27-04-2017]; Available from: <http://alias-i.com/lingpipe/demos/tutorial/read-me.html>.
26. Sharma, N.R. and V.D. Chitre, *Opinion Mining, Analysis and its Challenges*. International Journal of Innovations & Advancement in Computer Science IJIACS, 2014. **3**(1).
27. ChandraKala, S. and C. Sindhu, *Opinion Mining and Sentiment Classification: A Survey*. ICTACT Journal On Soft Computing, 2012. **3**(1).
28. Seerat, B. and F. Azam, *Opinion Mining: Issues and Challenges (A survey)*. International Journal of Computer Applications, 2012. **49**(9).
29. Blei, D., A. Ng, and M. Jordan, *Latent dirichlet allocation*. The Journal of Machine Learning Research, 2003.
30. Popescu, A.-M., B. Nguyen, and O. Etzioni, *OPINE: Extracting Product Features and Opinions from Reviews*, in *HLT/EMNLP 2005 Demonstration Abstracts*. 2005 Vancouver.
31. Qiu, G., et al., *Expanding domain sentiment lexicon through double propagation*, in *21st International Joint Conference on Artificial Intelligence*. 2009.
32. Agrawal, R. and R. Srikant, *Fast algorithms for mining association rules in large databases*, in *20th International Conference on Very Large Data Bases*. 1994.
33. Hu, M. and B. Liu, *Mining and summarizing customer reviews*, in *10th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2004: Seattle, USA.
34. Wang, B. and M. Liu, *Deep Learning for Aspect-Based Sentiment Analysis*. 2015.
35. Kim, Y., *Convolutional Neural Networks for Sentence Classification*, in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014.
36. Zhao, W.X., et al., *Jointly modeling aspects and opinions with a maxent-lda hybrid*, in *Conference on Empirical Methods in Natural Language Processing*. 2010, Association for Computational Linguistics.
37. Chen, H.-Y. and H.-H. Chen, *Implicit Polarity and Implicit Aspect Recognition in Opinion Mining*, in *54th Annual Meeting of the Association for Computational Linguistics*. 2016, Association for Computational Linguistics: Berlin, Germany.
38. Pereira, F., J. Lafferty, and A. McCallum, *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*. 2001.
39. Gindl, S., A. Weichselbraun, and A. Scharl, *Rule-based opinion target and aspect extraction to acquire affective knowledge*, in *22nd International Conference on World Wide Web*. 2013.
40. Poria, S., et al., *Sentic LDA: Improving on LDA with Semantic Similarity for Aspect-Based Sentiment Analysis*. 2016.
41. Arun, R., et al., *On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations*, in *Advances in Knowledge Discovery and Data Mining: 14th Pacific-Asia Conference (PAKDD)*. 2010: Hyderabad, India.
42. Brody, S. and N. Elhadad, *An unsupervised aspect-sentiment model for online reviews*. ACL:HLT, 2010.
43. Titov, I. and R. McDonald, *Modeling online reviews with multi-grain topic models*, in *17th international conference on World Wide Web*. 2008, ACM.
44. Vo, A.-D. and C.-Y. Ock *Sentiment Classification: A Combination of PMI, SentiWordNet and Fuzzy Function*. 2012
45. Su, Q., et al. *Using Pointwise Mutual Information to Identify Implicit Features in Customer Reviews*. 2006.
46. Zhang, L., et al., *Extracting and ranking product features in opinion documents*, in *23rd International Conference on Computational Linguistics*. 2010.
47. Liu, B., M. Hu, and J. Cheng, *Opinion Observer: Analyzing and Comparing Opinions*, in *14th International Conference on World Wide Web*. 2005.
48. Bengio, Y., A. Courville, and P. Vincent, *Unsupervised Feature Learning and Deep Learning: A Review and New Perspectives*. 2012 Department of computer science and operations research, University of Montreal.
49. Poria, S., E. Cambria, and A. Gelbukh, *Aspect extraction for opinion mining with a deep convolutional neural network*. Knowledge-Based Systems 2016. **108**.
50. Qiu, G., et al., *Opinion word expansion and target extraction through double propagation*. Computational Linguistics, 2011. **37**.

51. Liu, P., S. Joty, and H. Meng, *Fine-grained opinion mining with recurrent neural networks and word embeddings*, in *Conference on Empirical Methods in Natural Language Processing*. 2015.
52. Schmidhuber, J., *Deep learning in neural networks: An overview*, in *Neural Networks*. 2015, Elsevier Ltd.
53. Kessler, J.S. and N. Nicolov, *Targeting sentiment expressions through supervised ranking of linguistic configurations*, in *Third International AAAI Conference on Weblogs and Social Media*. 2009: San Jose, California, USA.
54. Singh, A. and M.E. Ullah, *Aspect based Sentiment Analysis*. 2015.
55. Malhotra, Nishtha, and et al., *SAP-RI: A Constrained and Supervised Approach for Aspect-Based Sentiment Analysis*, in *SemEval 2014*. 2014.
56. Brychcin, T., M. Konkol, and J. Steinberger, *UWB: Machine Learning Approach to Aspect-Based Sentiment Analysis*, in *SemEval 2014*. 2014: Dublin, Ireland.
57. Chernyshevich, M., *IHS R&D Belarus: Cross-domain Extraction of Product Features using Conditional Random Fields*, in *8th International Workshop on Semantic Evaluation (SemEval 2014)*. 2014: Dublin, Ireland.
58. Toh, Z. and W. Wang, *DLIREC: Aspect Term Extraction and Term Polarity Classification System*, in *8th International Workshop on Semantic Evaluation (SemEval 2014)*. 2014: Dublin, Ireland.
59. Ding, X., B. Liu, and L. Zhang, *Entity discovery and assignment for opinion mining applications*, in *15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2009.
60. P. Sanju and T.T. Mirnalinee, *Intelligent Robot with Aspect based Sentiment Analysis*. Téc. Ing. Univ. Zulia, 2016. **39**.
61. McCallum, A. and K. Nigam, *A Comparison of Event Models for Naive Bayes Text Classification*. AAAI Workshop on Learning for Text Categorization, 1998.
62. Jin, W., H.H. Ho, and R.K. Srihari, *OpinionMiner: a novel machine learning system for web opinion mining and extraction*, in *15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2009: Paris, France.
63. Gupta, D.K. and A. Ekbal, *IITP: Supervised Machine Learning for Aspect based Sentiment Analysis*, in *SemEval 2014*. 2014: Dublin, Ireland.
64. Jebbara, S. and P. Cimiano, *Aspect-Based Sentiment Analysis Using a Two-Step Neural Network Architecture*. 2016.
65. Clavel Quintero, Y., *Sistema de Clasificación Automática de Noticias a publicar en el periódico ¡ahora! digital*. 2010, Universidad Oscar Lucero Moya: Holguín.
66. Pang, B., L. Lee, and S. Vaithyanathan, *Sentiment classification using machine learning techniques*, in *ACL 2002 Conference on Empirical Methods in Natural Language Processing*. 2002, Association for Computational Linguistics.
67. Wagner, o., et al., *DCU: Aspect-based Polarity Classification for SemEval Task 4*, in *8th International Workshop on Semantic Evaluation (SemEval 2014)*. 2014: Dublin, Ireland.
68. Schouten, K., F. Frasincar, and F.d. Jong, *Commit-p1wp3: A co-occurrence based approach to aspect-level sentiment analysis*, in *8th International Workshop on Semantic Evaluation (SemEval 2014)*. 2014: Dublin, Ireland.
69. Cambria, E., D. Olsher, and D. Rajagopal, *SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis*. in *Twenty-Eighth AAAI Conference on Artificial Intelligence*. 2014.
70. Guha, S., A. Joshi, and V. Varma, *SIEL: Aspect Based Sentiment Analysis in Reviews*, in *SemEval-2015*. 2015.
71. Kiritchenko, S., et al., *NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews*, in *8th International Workshop on Semantic Evaluation (SemEval 2014)*. 2014: Dublin, Ireland.

7. SÍNTESIS CURRICULAR DEL AUTOR

Yisel Clavel Quintero nació en Baracoa, Guantánamo, el 1 de mayo de 1987. Cursó la carrera Ingeniería Informática en la Universidad de Holguín Oscar Lucero Moya, graduándose en el año 2010. En la misma Universidad cursó la maestría Matemática Aplicada e Informática para la Administración que defendió en el año 2015. Trabaja como profesora Asistente del Departamento de Informática de la Universidad de Holguín y dirige el Grupo Científico Estudiantil "Sistema de Gestión de la Universidad" y pertenece al proyecto "La gestión del conocimiento mediante técnicas de inteligencia artificial". Pertenecer a la Unión de Informáticos de Cuba. Líneas de trabajo: Sistemas de gestión de información, Sistemas de apoyo a la toma de decisiones, Inteligencia de Negocios, *data warehousing*, minería de datos, minería de texto, aprendizaje automático, minería de opinión.

Leticia Arco García nació el 1 de junio de 1978 en Santa Clara, Cuba. Se graduó de licenciatura y de máster en Ciencia de la Computación en la UCLV en el 2001 y en el 2005, respectivamente. En el 2009 obtuvo el grado de doctora en Ciencias Técnicas. En el 2015 concluyó una estancia post-doctoral en la Universidad Libre de Bruselas, Bélgica. Actualmente es profesora titular del departamento de Ciencia de la Computación y miembro del grupo de Inteligencia Artificial del CII-UCLV. Es asociada joven de la Academia de Ciencias de Cuba en el período 2012-2020. Es miembro de la Asociación Cubana de Reconocimiento de Patrones y la Sociedad Cubana de Matemática y Computación. Es miembro del tribunal nacional de Matemática y Computación. Está interesada en la minería de textos y de opinión, el aprendizaje automático y la inteligencia artificial.