



<https://creativecommons.org/licenses/by/4.0/>

MINERÍA DE OPINIÓN, ACCESO ABIERTO Y GOBERNANZA

Opinion mining, open access and governance

HECTOR NIGRO¹

Recibido: 20 de marzo de 2020. Aceptado: 15 de mayo de 2020

DOI: <http://dx.doi.org/10.21017/rimci.2020.v7.n14.a82>

RESUMEN

El objetivo de este artículo es presentar una nueva mirada de necesidades de investigación sobre minería de opiniones, acceso abierto de la información y gobernanza, donde el concepto de minería de opinión se hace cada vez más complejo y completo por las técnicas y tecnologías asociadas, sin embargo mientras la TI avanza, por otro lado, la formulación de políticas, leyes, normas y/o regulaciones decae así como el impacto del uso de la información en este mundo globalizado y polarizado.

Palabras clave. Minería de opinión; Comentarios; Información; Acceso Abierto; Gobernanza.

ABSTRACT

The aim of this paper is to present a new look at the research needs on opinion mining, open access to information and governance, where the concept of opinion mining becomes increasingly complex and complete due to the associated techniques and technologies, without However, as IT advances, on the other hand, the formulation of policies, laws, norms and / or regulations declines as well as the impact of the use of information in this globalized and polarized world.

Keywords. Opinion mining; Comments; Information; Open access; Governance.

I. INTRODUCCIÓN

EL AUGE de las redes sociales ha creado oportunidades sin precedentes para que las personas expresen públicamente sus opiniones, detonando un sinnúmero de situaciones de orden público, religioso, político e incluso económico. Por esta razón se requiere un control de opiniones en las redes para poder encontrar sentido y ejercer un monitoreo sobre este tipo de comentarios e intereses. En la actualidad, ha aumentado la necesidad de obtener una comprensión en tiempo real de los comentarios en las redes: debido a la naturaleza viral de estos espacios virtuales (donde la atención se distribuye de manera muy desigual y rápida).

Es así como, los gobiernos aún no tienen una forma efectiva de dar sentido conversaciones masivas, el debate público en las redes sociales se

caracteriza por el cortoplacismo y autorreferencialidad. Muchos expertos consideran que las redes sociales son una oportunidad perdida para un mejor debate sobre políticas y, al mismo tiempo, la gran cantidad de datos sin procesar también es una oportunidad para entender mejor las opiniones.

Al realizar esta segmentación existen múltiples beneficios para las empresas, los gobiernos, etc., al comprender lo que el público piensa acerca de sus productos y servicios, pero también le interesa a las grandes instituciones de conocimiento público poder recopilar, recuperar y preservar toda la información relacionada con ciertos eventos y su desarrollo. Con el tiempo, la difusión de información a través de las redes sociales también puede desencadenar una cadena de reacciones a tales situaciones y eventos que

¹ Ingeniero de Sistemas (Unicen), Magister en Ciencias Políticas y Sociales (Flacso), Candidato a Doctor en Matemática Computacional e Industrial Aplicada (Unicen). Docente Investigador. Correo electrónico: oscarnigro@unicer.com.ar

finalmente conducen a cambios administrativos, políticos y sociales.

II. MINERÍA DE OPINIONES

El desafío de la sobrecarga de información puede convertirse no en un problema, sino en una oportunidad para dar sentido a mil voces e identificar problemas tan pronto como surjan [1]. La minería de opiniones se puede definir como una subdisciplina de la lingüística computacional que se centra en extraer la opinión de las personas de la web.

Es así como, la reciente expansión de la web anima a los usuarios a contribuir y expresarse a través de blogs, videos, sitios de redes sociales, etc. Todas estas plataformas brindan una enorme cantidad de información valiosa que es de interés para analizar.

Dado un fragmento de texto, los sistemas de minería de opinión analizan: ¿Qué parte es la opinión que expresa?, ¿Quién escribió la opinión?, ¿Qué se está comentando?, además puede proporcionar el análisis de sentimiento, que se trata de determinar la subjetividad, polaridad (positiva o negativa) y la fuerza de la polaridad (débilmente positiva, medianamente positiva, fuertemente positiva, etc.) de un texto; en otras palabras: ¿Cuál es la opinión y actitud del escritor frente a un determinado tema? [2].

III. MINERÍA DE OPINIÓN Y ACCESO ABIERTO A LA INFORMACIÓN

El uso masivo exponencial de redes sociales en línea y el alcance progresivo y omnipresente de la TI en todos los dispositivos ha detonado una tendencia de acceso abierto, que se caracteriza por el hecho de que en la gran mayoría de países, los gobiernos publican cada vez más sus repositorios de datos para que otras personas accedan a ellos y los utilicen.

Otra tendencia se refiere a la gran cantidad de las personas ponen a disposición los datos a través de la detección participativa, donde las personas toman un papel proactivo en la publicación de comentarios y quejas en línea, y utilizan cada vez más

la tecnología para registrar información adicional, como fotos o grabaciones de audio, generalmente a través de teléfonos inteligentes.

De este modo que los datos utilizables se recopilan automáticamente a un ritmo muy rápido y además de que son recolectados, se ponen a disposición de los ciudadanos para que haya nueva información disponible, así se forma un ciclo de información garantizando la sostenibilidad de datos en la red.

Es aquí donde, a nivel analítico, existen varias innovaciones tecnológicas en minería de opiniones que ayudan a dar sentido a la gran cantidad de datos disponibles. En este sentido, se trata de un desarrollo específico de una disciplina que se ocupa de la extracción de información no estructurada (IE) que anteriormente trabajaba principalmente con datos objetivos como desastres naturales o información bibliográfica. La explosión del contenido generado por los usuarios amplía el ámbito de aplicación de las herramientas de minería de opinión pública, que se están volviendo más omnipresentes y disponibles para la mayoría de las personas.

A. Aplicaciones de la minería de opiniones

El software de mapeo de argumentos ayuda a organizar de manera lógica la semántica de las opiniones, al explicar los vínculos lógicos entre ellas. En el campo de investigación de la deliberación en línea, se han desarrollado herramientas como Compendium, Debatepedia, Cohere, Debategraph para dar una estructura lógica a una serie de declaraciones y para vincular argumentos con la evidencia que la respalde.

En el ámbito político esto es muy funcional, porque los votantes entienden qué partido político (u otros votantes) tienen posiciones más cercanas a las suyas. Por ejemplo, SmartVote.ch le pide al votante que declare su grado de acuerdo con una serie de declaraciones de política, luego compara su posición con los partidos políticos.

El análisis de contenido automatizado ayuda a procesar una gran cantidad de datos cualitativos. Hoy en día existen en el mercado muchas herramientas que combinan algoritmos estadísticos con semántica y ontologías, así como aprendizaje

automático con supervisión humana. Estas soluciones son capaces de identificar comentarios relevantes y asignarles connotaciones positivas o negativas (el llamado sentimiento).

Con base en lo anterior, algunos aspectos reflejan áreas de aplicación que ya se han venido trabajando desde tiempo atrás, pero otras están emergiendo.

IV. GOBERNANZA

En la actualidad, las aplicaciones de minería de opiniones son la infraestructura básica de la formulación de políticas colaborativas a gran escala. Ayudan a dar sentido a miles de intervenciones. Colaboran con la detección del sistema de alerta temprana de una posible interrupción de manera oportuna, al detectar la retroalimentación temprana de los ciudadanos. Tradicionalmente, las encuestas ad hoc se utilizan para recopilar comentarios de manera estructurada. Sin embargo, este tipo de recopilación de datos es costosa, ya que merece una inversión en diseño y recopilación de datos; es difícil, ya que la gente no está interesada en responder encuestas; y finalmente no es muy valioso, ya que detecta “problemas conocidos” a través de preguntas predefinidas y entrevistas, pero no detecta los problemas más importantes, el famoso “desconocido desconocido”.

La minería de opiniones es útil para identificar problemas escuchando, en lugar de preguntar, asegurando así un reflejo más preciso de la realidad. El software de mapeo de argumentos resulta entonces útil para garantizar que los debates políticos sean lógicos y estén basados en evidencias, y no repitan los mismos argumentos una y otra vez. Estas herramientas finalmente serían útiles no solo para los responsables políticos, sino también para los ciudadanos que podrían hacerlo más fácilmente, comprender los puntos clave de una discusión y participar en el proceso de formulación de políticas.

La acción de los gobiernos en la minería de opiniones no es reciente, sus intereses particulares datan desde el inicio de la sociedad de la información, es así como los métodos automatizados para el análisis de contenido se han utilizado cada vez más y han aumentado exponencialmente [3].

Además de técnicas de procesamiento del lenguaje natural, la minería de textos, el aprendizaje automático y la inteligencia artificial, el análisis de contenido automatizado y las aplicaciones de asesoramiento para la votación (desde el ámbito político).

En ese sentido, cobra valor el desarrollo tecnológico pero disminuye cada vez más los derechos de los ciudadanos, regulaciones frente al uso de información y mecanismos de control.

El contenido de las redes sociales refleja por naturaleza opiniones y sentimientos, mientras que el de contenido tradicional tendía a centrarse en identificar temas [4]. Como tal, se ocupa de problemas más complejos del lenguaje natural. Debido a la combinación de aumento de el volumen de datos disponibles y los conceptos más complejos para analizar, en los últimos años ha disminuido el interés por la aplicación de base semántica y se ha avanzado hacia un mayor uso de la estadística y la visualización.

Al igual que cualquier otra disciplina científica, también el análisis de contenido automatizado se está convirtiendo en una ciencia intensiva en datos.

V. HERRAMIENTAS

El mercado de herramientas de minería de opinión está lleno de proveedores de soluciones. La mayoría de estas aplicaciones están orientadas a analizar los comentarios de los clientes sobre productos y servicios y, por lo tanto, se inclinan hacia el análisis de sentimientos que detecta sentimientos positivos / negativos mediante la interpretación del lenguaje natural. Herramientas de libre disponibilidad, la mayoría de estas, son de asesoramiento al votante y mapeo de argumentos de última generación están disponibles gratuitamente, ya que derivan en gran parte de la comunidad académica o de ONG.

Por otra parte, existe el flujo de soluciones simples, gratuitas y populares con la visualización de palabras. Las nubes de palabras se utilizan cada vez más para dar sentido a grandes cantidades de información de una forma instantánea. Estas herramientas también están divididas y solo ofrecen una visualización de los términos más comunes, lo

que es útil para tener una idea de lo que trata el documento, comentario o fragmento de texto a analizar.

Herramientas como wordle.com proporcionan una solución de diseño atractiva que puede servir como un nivel de entrada en el mercado de la minería de opiniones. Por tanto, son importantes para involucrar a un público mucho más amplio en este tipo de actividades.

Finalmente, otra forma de dar sentido a una gran cantidad de información es apoyándose en el esfuerzo humano, mediante el **crowdsourcing** y la **inteligencia colectiva**: las personas no solo están enviando sus opiniones, sino que en realidad las filtran señalando las más importantes. Herramientas como uservoice.com permiten a los clientes enviar comentarios y clasificar las ideas de otras personas, lo que permite que surjan las ideas más populares.

Estas herramientas están disponibles a muy bajo costo, pero las investigaciones muestran que son efectivas para recopilar comentarios pero no para identificar buenas ideas, ya que la votación tiende a enfocarse en temas más fáciles y populares.

Además de estas aplicaciones simples y gratuitas, existe un mercado de software de nivel empresarial para la extracción de opiniones con características mucho más avanzadas. Estas herramientas son utilizadas en gran medida por las empresas para monitorear su reputación y los comentarios sobre los productos en las redes sociales. En el contexto gubernamental, la minería de opinión se ha utilizado durante mucho tiempo como una herramienta de inteligencia para detectar comunicaciones hostiles o negativas [5].

Recientemente, la política se ha convertido en un área clave de aplicaciones, ya que los políticos monitorean la opinión pública en las redes sociales para comprender la reacción del público a su posición.

Técnicamente, estas herramientas se basan en el aprendizaje automático para identificar y clasificar los comentarios relevantes, a través de una combinación de elementos latentes, análisis semántico, máquinas de vectores de soporte y orientación semántica. Este proceso requiere un

esfuerzo humano significativo con la ayuda de máquinas: todas las herramientas en el mercado se basan en una combinación de análisis de máquina y humano, generalmente utilizando máquinas para aumentar la capacidad humana para clasificar, codificar y etiquetar comentarios. El análisis automatizado se basa en una combinación de análisis semántico y estadístico. En ese orden de ideas, debido al gran aumento en la cantidad de conjuntos de datos disponibles, el análisis estadístico se está volviendo más importante.

VI. CONCLUSIONES

Las soluciones actuales para la minería de opiniones y el análisis de sentimientos están evolucionando rápidamente, típicamente reduciendo la cantidad de esfuerzo humano necesario para clasificar los comentarios. Entre los desafíos identificados se puede destacar: la detección de spam y reseñas falsas, principalmente a través de la identificación de duplicados, la comparación de revisiones cualitativas con reseñas resumidas, la detección de valores atípicos y la reputación del revisor [6]: los límites del filtrado colaborativo, que tiende a identificar los conceptos más populares y a pasar por alto, el pensamiento más innovador, donde el análisis de contenido automatizado combinado con el análisis de comportamiento conduce a una selección muy efectiva pero en última instancia desviada de opiniones y contenido relevantes, de modo que el usuario no es consciente de contenido que de alguna manera es diferente de sus expectativas: la asimetría en la disponibilidad del software de minería de opiniones, que actualmente solo las organizaciones pueden ofrecer y el gobierno pero no por los ciudadanos precisamente sino para sus propios intereses, la pregunta es ¿cómo regular este tipo de tecnologías?, útiles, válidas pero que se pueden salir de control, no está mal que el gobierno tenga los medios hoy para monitorear la opinión pública en formas que no están disponibles para el ciudadano promedio. Si bien la producción y publicación de contenido se ha democratizado, el análisis de contenido no. La integración de la opinión con el comportamiento y los datos implícitos, para validar y proporcionar un análisis más profundo de los datos más allá de la opinión expresada y la necesidad continua de una mejor usabilidad y facilidad de uso de las herramientas, que actualmente son utilizadas principalmente por analistas de datos.

Así las cosas, técnicamente se requiere mejorar la precisión de algoritmos para la detección de opiniones, reducir el esfuerzo humano necesario para analizar el contenido, mejorar el análisis semántico a través del léxico / corpus de palabras con sentimiento conocido para la clasificación del sentimiento, identificar el material de opinión sobre políticas que se analizarán, mejorar mapeo visual de la opinión bipolar, entre otros aspectos.

Es una realidad que cada vez se tendrá mayor capacidad de descubrimiento de contenido a través de datos vinculados, Representación visual con mayor facilidad de lectura de datos, minería de opiniones audiovisuales, minería de opiniones en tiempo real, algoritmos de aprendizaje automático, evaluación bipolar de opiniones, minería de opinión multiplataforma, aprendizaje automático e inteligencia artificial autónomos y con la gobernaza esperada de cada una de las naciones se espera tener herramientas de minería de opinión de igual a igual para todos los ciudadanos.

REFERENCIAS

- [1] D. Osimo and F. Mureddu, Desafío de investigación sobre minería de opiniones y análisis de sentimientos. 2012.
- [2] E. Boiy, Pieter Hens, Koen Deschacht, and Marie-Francine Moens. Automatic sentiment analysis of onlinetext. InProc. of the 11th International Conference on Electronic Publishing, Vienna, Austria. 2007.
- [3] K. A. Neuendorf, The ContentAnalysis Guidebook. Sage. 2002.
- [4] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. Information Retrieval, 2(1). 2008.
- [5] A. Scharl and A. Weichselbraun. An automated approach to investigating the online media coverage of USpresidential elections. Journal of Information Technology and Politics, 5(1):121-132. 2008.
- [6] Johan Bollen and Huina Mao. Twitter mood as a stock market predictor. IEEE Computer, 44(10):91-94. 2011.

